



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address : COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, VA 22313-1450
www.uspto.gov

APPLICATION NO.	ISSUE DATE	PATENT NO.
16/707,265	16-MAY-23	11650968

Pearl Cohen Zedek Latzer Baratz LLP
7 Times Square, 19th Floor
New York, NY 10036

EGRANT NOTIFICATION

Your electronic patent grant (eGrant) is now available, which can be accessed via Patent Center at <https://patentcenter.uspto.gov>

The electronic patent grant is the official patent grant under 35 U.S.C. 153. For more information, please visit <https://www.uspto.gov/electronicgrants>



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	ISSUE DATE	PATENT NO.	ATTORNEY DOCKET NO.	CONFIRMATION NO.
16/707,265	05/16/2023	11650968	P-587424-US	1015

49443 7590 04/26/2023

Pearl Cohen Zedek Latzer Baratz LLP
7 Times Square, 19th Floor
New York, NY 10036

ISSUE NOTIFICATION

The projected patent number and issue date are specified above. The patent will issue electronically. The electronically issued patent is the official patent grant pursuant to 35 U.S.C. § 153. The patent may be accessed on or after the issue date through Patent Center at <https://patentcenter.uspto.gov/>. The patent will be available in both the public and the private sides of Patent Center. Further assistance in electronically accessing the patent, or about Patent Center, is available by calling the Patent Electronic Business Center at 1-888-217-9197.

The USPTO is implementing electronic patent issuance with a transition period, during which period the USPTO will mail a ceremonial paper copy of the electronic patent grant to the correspondence address of record. Additional copies of the patent (i.e., certified and presentation copies) may be ordered for a fee from the USPTO's Certified Copy Center at <https://certifiedcopycenter.uspto.gov/index.html>. The Certified Copy Center may be reached at (800)972-6382.

Determination of Patent Term Adjustment under 35 U.S.C. 154 (b) (application filed on or after May 29, 2000)

The Patent Term Adjustment is 683 day(s). Any patent to issue from the above-identified application will include an indication of the adjustment on the front page.

If a Continued Prosecution Application (CPA) was filed in the above-identified application, the filing date that determines Patent Term Adjustment is the filing date of the most recent CPA.

Applicant will be able to obtain more detailed information by accessing the Patent Application Information Retrieval (PAIR) WEB site (<http://pair.uspto.gov>).

Any questions regarding the Patent Term Extension or Adjustment determination should be directed to the Office of Patent Legal Administration at (571)-272-7702. Questions relating to issue and publication fee payments should be directed to the Application Assistance Unit (AAU) of the Office of Patents Stakeholder Experience (OPSE), Stakeholder Support Division (SSD) at (571)-272-4200.

INVENTOR(s) (Please see PAIR WEB site <http://pair.uspto.gov> for additional inventors):

Dhruv NAIR, New York, NY;
Gideon MENDELS, New York, NY;
Nimrod LAHAV, Hoboken, NJ;

APPLICANT(s) (Please see PAIR WEB site <http://pair.uspto.gov> for additional applicants):

Comet ML, Inc., New York, NY;

The United States represents the largest, most dynamic marketplace in the world and is an unparalleled location for business investment, innovation, and commercialization of new technologies. The USA offers tremendous resources and advantages for those who invest and manufacture goods here. Through SelectUSA, our nation works to encourage and facilitate business investment. To learn more about why the USA is the best country in the world to develop technology, manufacture products, and grow your business, visit SelectUSA.gov.



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
16/707,265	12/09/2019	Dhruv NAIR	P-587424-US	1015
49443	7590	04/07/2023		
Pearl Cohen Zedek Latzer Baratz LLP			EXAMINER	
7 Times Square, 19th Floor			ALLISON, ANDRAE S	
New York, NY 10036				
			ART UNIT	PAPER NUMBER
			2663	
			NOTIFICATION DATE	DELIVERY MODE
			04/07/2023	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

Arch-USPTO@PearlCohen.com
USPTO@PearlCohen.com

Response to Rule 312 Communication	Application No. 16/707,265	Applicant(s) NAIR et al.	
	Examiner ANDRAE S ALLISON	Art Unit 2663	AIA (FITF) Status Yes
-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address -- 1. ☒ The amendment filed on <u>29 March 2023</u> under 37 CFR 1.312 has been considered, and has been: a) ☐ entered. b) ☒ entered as directed to matters of form not affecting the scope of the invention. c) ☐ disapproved because the amendment was filed after the payment of the issue fee. Any amendment filed after the date the issue fee is paid must be accompanied by a petition under 37 CFR 1.313(c)(1) and the required fee to withdraw the application from issue. d) ☐ disapproved. See explanation below. e) ☐ entered in part. See explanation below. f) ☐ not entered because the supplemental or corrected Application Data sheet (ADS) ☐ was not accompanied by a petition to accept an unintentionally delayed claim under 37 CFR 1.55 or 27 CFR 1.78; ☐ did not identify the information being changed in accordance with 37 CFR 1.76(c)(2); ☐ was not properly signed in accordance with 37 CFR 1.76(e) (or 37 CFR 1.33(b) for applications filed prior to September 16, 2012). _____			
/ANDRAE S ALLISON/ Primary Examiner, Art Unit 2663			

OK TO ENTER: /A.S.A/

Attorney Docket No.: P-587424-US

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Inventor(s):	NAIR, Dhruv et al.	Examiner:	ALLISON, Andrae S
Serial No.:	16/707,265	Group Art Unit:	2663
Filed:	December 9, 2019	Confirmation No.:	1015
Title:	SYSTEMS AND METHODS FOR PREDICTIVE EARLY STOPPING IN NEURAL NETWORK TRAINING		

AMENDMENT UNDER 37 CFR §1.312

Mail Stop Issue Fee
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir/Madam:

Pursuant to 37 CFR §1.312, kindly amend the above-identified application as follows:

Amendments to the Specification begin on page **2** of this Amendment.

Remarks/Arguments begin on page **6** of this paper.

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Inventor(s):	NAIR, Dhruv et al.	Examiner:	ALLISON, Andrae S
Serial No.:	16/707,265	Group Art Unit:	2663
Filed:	December 9, 2019	Confirmation No.:	1015
Title:	SYSTEMS AND METHODS FOR PREDICTIVE EARLY STOPPING IN NEURAL NETWORK TRAINING		

AMENDMENT UNDER 37 CFR §1.312

Mail Stop Issue Fee
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir/Madam:

Pursuant to 37 CFR §1.312, kindly amend the above-identified application as follows:

Amendments to the Specification begin on page **2** of this Amendment.

Remarks/Arguments begin on page **6** of this paper.

AMENDMENTS TO THE SPECIFICATION

In the Specification:

Please replace paragraph number [0014] with the following rewritten paragraph:

-- **[0014]** Systems and methods of the present invention may train one or more neural networks (NNs) and determine when to stop training to not waste computing or other resources when improvement is not no longer likely. Systems and methods may train neural networks (NNs) and determine when to stop training to not waste computing or other resources when improvement is not no longer likely. After training period (e.g. an epoch) for a NN, a model trained using training data from other NNs may return a [[a]] probability of improvement in the loss of the NN or a probability that the likely best loss of the NN is lower than the best loss of the other NNs for which hyperparameters have been chosen. Training may be stopped if the probability is less than a threshold, or a wait value is greater than a wait threshold.--

Please replace paragraph number [0045] with the following rewritten paragraph:

-- **[0045]** In one embodiment, a final model may include a set of individual models which may be responsible for predicting the minimum value of convergence based on the current training epoch, such as the following example:

$$f(x) = \{\{m_1^{10}(x^{10}), m_2^{10}(x^{10}), m_3^{10}(x^{10})\}, \dots, \{m_1^t(x^t), m_2^t(x^t), m_3^t(x^t)\}\}$$

Where m_1^X is a first model, returning an estimated mean, for epoch or other period of time X (and which may be considered to have X sub-models each corresponding to an epoch or other period of time), m_2^X is a second model, returning an upper or high bound for an estimated range (a quantile loss that estimates an upper percentile), for epoch X; and m_3^X is a third model, returning a lower or low bound for an estimated loss (a quantile loss that estimates a lower percentile), for epoch X. In one embodiment, a wait period is applied such that the model is not applied until the Xth (e.g. 10th) epoch is used, as typically loss will decrease during initial training epochs and premature stopping due to actual loss being greater than

estimated loss is not desired during this early period. In such a model each leaf describes NNs having been trained on the same number of epochs.--

Please replace paragraph number [0049] with the following rewritten paragraph:

-- [0049] After preprocessing (e.g. normalizing, smoothing, etc.) certain features of the values may be extracted or calculated and used in models to categorize instances of NN behavior, e.g. to categorize NN instances as leaves in a tree model. The features extracted may be used to decide into which leaf or category to place the NN. Extracted time series features may be input along with hyperparameters and graph features to an alignment process. Hyperparameters and other NN characteristics may be extracted from for example a graph representing the NN (a NN may be represented as a graph of nodes connected by links). Extracted timeseries features, hyperparameters (e.g. batch size, learning rate, number of training samples, etc.), and graph (e.g. extracted from the graph representing the NN) or other features such as the number of ~~trainable~~ trainable parameters present in the mode may be analyzed. Example values or features that can be extracted to describe training loss include:

- First and second order difference of the normalized or pre-processed time series;
- Summary statistics of the first and second order differential features, such as the mean, median, sum, standard deviation, etc.
- Autocorrelation coefficients of the raw, untransformed time series data. The mean and standard deviation of this set of data may be used as features. Autocorrelation coefficients of a time series may provide a measure of how strongly a value in the time series correlates to a lagged version of itself. For example, when a lag value of 1 is used, how strongly correlated the time series at t-1 is to the time series at t may be calculated. A lag of 1 will result in a single autocorrelation coefficient. In some embodiments, a lag of, for example, up to 40 may be used, resulting in a maximum of 40 coefficients. The mean and standard deviation of these coefficients may be used as features.

The raw time series features used to create the derived data may be dropped or ignored, or may be set to "not a number", or NAN. Other or different features may be used.--

Please replace paragraph number [0058] with the following rewritten paragraph:

-- **[0058]** Computer 10 may communicate with learning server or computer 20, which may create, update and/or distribute early stopping module 14' or other software. Communication may be via network 30, e.g. the Internet. Computer 20 may communicate with computers 10 to receive from training software 14 training information or data 22, and to process and store data 22 in order to update and/or create stopping module 14'. Computer [[22]] 20 may distribute stopping module 14' or updates to stopping module 14' to computers 10. While in the example shown updates to stopping module 14' are performed by a computer different from a computer using stopping module 14' for training, in other embodiments a computer conducting training may also update or create a stopping module. Other different configurations are possible: for example a computer 10 controlling training and interfacing with a user may outsource some or all of training (including a stopping module 14') to a remote or cloud computer. While stopping module 14' may when executed by a computer 10 indicate when NN training should stop (e.g. by returning True, indicating stop or False, indicating do not stop; other return values indicating stopping or not stopping may be used), in other embodiments the functionality of stopping module may be performed by computer 20 communicating with a computer 10 conducting training, or the functionality of stopping module may be performed by a different remote, e.g. cloud, computer. Examples of functionality stopping module 14' are shown in Tables 1-5 herein.--

Please replace paragraph number [0095] with the following rewritten paragraph:

-- **[0095]** The graph in Fig. 7B shows the results of early stopping algorithm with various threshold and patience values. In Fig. 7B the Y axis is the difference or error between the change in loss achieved during full training (e.g. without early stopping, rather [[the]] when training algorithm stops naturally according to prior art methods) and the change in loss at the predicted stop point. In Fig. 7B, if the actual change to minimum using prior art methods was 97% and an early stopping algorithm stopped training at 96%, the error will be 1% or 0.01. The X axis is the value of the patience input control parameter. Line 810 (orange) shows the error or difference input accuracy threshold value of 0.01, line 812 (blue) shows error or

APPLICANT(S): NAIR, Dhruv et al.
SERIAL NO.: 16/707,265
FILED: December 9, 2019
Page 5

difference for an input accuracy threshold value of 0.05, and line 814 (green) shows error or difference for an input accuracy threshold value of 0.1. Fig. 7B shows that using one embodiment, error may follow the threshold value quite closely.--

APPLICANT(S): NAIR, Dhruv et al.
SERIAL NO.: 16/707,265
FILED: December 9, 2019
Page 6

REMARKS

The foregoing amendments and remarks are directed to formalities, in order to move the above-referenced patent application to issue. Entry of this paper is respectfully requested.

The amendments to the specification are editorial in nature and are meant to correct form and language only.

It is respectfully asserted that none of the above listed changes add new matter. Approval is requested to the above listed changes.

Should the Examiner have any question or comment as to the form, content or entry of this Amendment, the Examiner is requested to contact the undersigned at the telephone number below. Similarly, if there are any further issues yet to be resolved to advance the prosecution of this application to issue, the Examiner is requested to telephone the undersigned counsel.

To the extent any fees are due, those fees are being paid by credit card and/or deposit account. However, to the extent the fees due are not specifically paid, please charge any fees associated with this paper to deposit account No. 50-3355.

Respectfully submitted,

/Caleb Pollack/

Caleb Pollack
Attorney/Agent for Applicant(s)
Registration No. 37,912

Dated: March 29, 2023

Pearl Cohen Zedek Latzer Baratz LLP
1500 Broadway, 12th Floor
New York, New York 10036
Tel: (646) 878-0800
Fax: (646) 878-0801

PART B – FEE(S) TRANSMITTAL

Complete and send this form, together with applicable fee(s), by mail or fax, or via EFS-Web.

By mail, send to: Mail Stop ISSUE FEE
Commissioner for Patents
P.O. Box 1450
Alexandria, Virginia 22313-1450

By fax, send to: (571) 273-2885

INSTRUCTIONS: This form should be used for transmitting the ISSUE FEE and PUBLICATION FEE (if required). Blocks 1 through 5 should be completed where appropriate. All further correspondence including the Patent, advance orders and notification of maintenance fees will be mailed to the current correspondence address as indicated unless corrected below or directed otherwise in Block 1, by (a) specifying a new correspondence address; and/or (b) indicating a separate "FEE ADDRESS" for maintenance fee notifications.

CURRENT CORRESPONDENCE ADDRESS (Note: Use Block 1 for any change of address)

49443 7590 02/17/2023

PEARL COHEN ZEDEK LATZER BARATZ LLP
7 TIMES SQUARE, 19TH FLOOR
NEW YORK, NY 10036

Note: A certificate of mailing can only be used for domestic mailings of the Fee(s) Transmittal. This certificate cannot be used for any other accompanying papers. Each additional paper, such as an assignment or formal drawing, must have its own certificate of mailing or transmission.

Certificate of Mailing or Transmission

I hereby certify that this Fee(s) Transmittal is being deposited with the United States Postal Service with sufficient postage for first class mail in an envelope addressed to the Mail Stop ISSUE FEE address above, or being facsimile transmitted to the USPTO (571) 273-2885, on the date indicated below.

(Typed or printed name)
(Signature)
(Date)

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
16/707,265	12-09-2019	Dhruv NAIR	P-587424-US	1015

TITLE OF INVENTION: SYSTEMS AND METHODS FOR PREDICTIVE EARLY STOPPING IN NEURAL NETWORK TRAINING

APPLN. TYPE	ENTITY STATUS	ISSUE FEE DUE	PUBLICATION FEE DUE	PREV. PAID ISSUE FEE	TOTAL FEE(S) DUE	DATE DUE
NONPROVISIONAL	SMALL	\$480	\$0.00	\$0.00	\$480	05/17/2023

EXAMINER	ART UNIT	CLASS-SUBCLASS
Andrae S. ALLISON	2663	706-020000

1. Change of correspondence address or indication of "Fee Address" (37 CFR 1.363).

- ☐ Change of correspondence address (or Change of Correspondence Address form PTO/SB/122) attached.
☐ "Fee Address" indication (or "Fee Address" Indication form PTO/SB/47; Rev 03-09 or more recent) attached. **Use of a Customer Number is required.**

2. For printing on the patent front page, list

- (1) the names of up to 3 registered patent attorneys or agents OR, alternatively,
(2) the name of a single firm (have as a member a registered attorney or agent) and the names of up to 2 registered patent attorneys or agents. If no name is listed, no name will be printed.

1 PEARL COHEN ZEDEK

2 LATZER BARATZ LLP

3 _____

3. ASSIGNEE NAME AND RESIDENCE DATA TO BE PRINTED ON THE PATENT (print or type)

PLEASE NOTE: Unless an assignee is identified below, no assignee data will appear on the patent. If an assignee is identified below, the document must have been previously recorded, or filed for recordation as set forth in 37 CFR 3.11 and 37 CFR 3.81(a). Completion of this form is NOT a substitute for filing an assignment.

(A) NAME OF ASSIGNEE

Comet ML, Inc.

(B) RESIDENCE: (CITY and STATE OR COUNTRY)

New York, USA

Please check the appropriate assignee category or categories (will not be printed on the patent): ☐ Individual ☒ Corporation or other private group entity ☐ Government

4a. Fees submitted: ☒ Issue Fee ☐ Publication Fee (if required) ☐ Advance Order - # of Copies _____

4b. Method of Payment: (Please first reapply any previously paid issue fee shown above)

☒ Electronic Payment via EFS-Web ☐ Enclosed check ☐ Non-electronic payment by credit card (Attach form PTO-2038)

☒ To the extent any fees due associated with this filing are not paid or not able to be paid by credit card, and to the extent any overpayment is made, the Director is hereby authorized to charge the required fee(s), or credit any overpayment to Deposit Account No. **50-3355**

5. Change in Entity Status (from status indicated above)

☐ Applicant certifying micro entity status. See 37 CFR 1.29

☐ Applicant asserting small entity status. See 37 CFR 1.27

☐ Applicant changing to regular undiscounted fee status.

NOTE: Absent a valid certification of Micro Entity Status (see form PTO/SB/15A and 15B), issue fee payment in the micro entity amount will not be accepted at the risk of application abandonment.

NOTE: If the application was previously under micro entity status, checking this box will be taken to be a notification of loss of entitlement to micro entity status.

NOTE: Checking this box will be taken to be a notification of loss of entitlement to small or micro entity status, as applicable.

NOTE: This form must be signed in accordance with 37 CFR 1.31 and 1.33. See CFR 1.4 for signature requirements and certifications.

Authorized Signature: _____/Caleb Pollack/_____

Date: March 29, 2023

Typed or printed name: Caleb Pollack

Registration No: 37,912

Page of

Electronic Patent Application Fee Transmittal				
Application Number:		16707265		
Filing Date:		09-Dec-2019		
Title of Invention:		SYSTEMS AND METHODS FOR PREDICTIVE EARLY STOPPING IN NEURAL NETWORK TRAINING		
First Named Inventor/Applicant Name:		Dhruv NAIR		
Filer:		Caleb Pollack/Keren Peretz		
Attorney Docket Number:		P-587424-US		
Filed as Small Entity				
Filing Fees for Utility under 35 USC 111(a)				
Description	Fee Code	Quantity	Amount	Sub-Total in USD(\$)
Basic Filing:				
Pages:				
Claims:				
Miscellaneous-Filing:				
Petition:				
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				
UTILITY APPL ISSUE FEE	2501	1	480	480

Description	Fee Code	Quantity	Amount	Sub-Total in USD(\$)
Extension-of-Time:				
Miscellaneous:				
Total in USD (\$)				480

Electronic Acknowledgement Receipt	
EFS ID:	47757351
Application Number:	16707265
International Application Number:	
Confirmation Number:	1015
Title of Invention:	SYSTEMS AND METHODS FOR PREDICTIVE EARLY STOPPING IN NEURAL NETWORK TRAINING
First Named Inventor/Applicant Name:	Dhruv NAIR
Customer Number:	49443
Filer:	Caleb Pollack/Keren Peretz
Filer Authorized By:	Caleb Pollack
Attorney Docket Number:	P-587424-US
Receipt Date:	29-MAR-2023
Filing Date:	09-DEC-2019
Time Stamp:	07:36:36
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	yes
Payment Type	CARD
Payment was successfully received in RAM	\$480
RAM confirmation Number	E20233S837316774
Deposit Account	503355
Authorized User	Keren Peretz
The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows: 37 CFR 1.16 (National application filing, search, and examination fees) 37 CFR 1.17 (Patent application and reexamination processing fees)	

37 CFR 1.19 (Document supply fees)					
37 CFR 1.20 (Post Issuance fees)					
37 CFR 1.21 (Miscellaneous fees and charges)					
File Listing:					
Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1		P-587424-US-U-312AMND-ISSFEE-29MAR23.pdf	114416	yes	9
			63ccb38f7c264f147964405e165bbd71b289527b		
	Multipart Description/PDF files in .zip description				
	Document Description		Start	End	
	Miscellaneous Incoming Letter		1	2	
	Amendment after Notice of Allowance (Rule 312)		3	3	
	Specification		4	7	
	Applicant Arguments/Remarks Made in an Amendment		8	8	
	Issue Fee Payment (PTO-85B)		9	9	
Warnings:					
Information:					
2	Fee Worksheet (SB06)	fee-info.pdf	38238	no	2
			c87ac7ba759fdd962b5925e5d53a25c671851a59		
Warnings:					
Information:					
Total Files Size (in bytes):			152654		

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Inventor(s):	Dhruv NAIR et al.	Examiner:	Andrae S. ALLISON
Serial No.:	16/707,265	Group Art Unit:	2663
Filed:	December 09, 2019	Confirmation No.:	1015
Title:	SYSTEMS AND METHODS FOR PREDICTIVE EARLY STOPPING IN NEURAL NETWORK TRAINING		

COMMENTS ON STATEMENT OF REASONS FOR ALLOWANCE

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir/Madam:

Applicant(s) would like to thank the Examiner for the allowance of the subject application based on the distinctions enumerated in the statement of reasons for allowance. Applicant(s) agree, without prejudice, that these distinctions provide sufficient basis for allowance of the application. However, Applicant(s) respectfully submit that there may be alternative and/or additional patentable distinctions between the claimed invention and the cited prior art, related or unrelated to those explicitly addressed in the statement of reasons for allowance.

Should the Examiner have any question or comment as to the form, content or entry of this paper, the Examiner is requested to contact the undersigned counsel at the telephone number below.

APPLICANT(S): Dhruv NAIR et al.
SERIAL NO.: 16/707,265
FILED: December 09, 2019
Page 2

To the extent that there are any fees associated with this filing that are due but are not paid or are not able to be paid by a credit card mentioned elsewhere in this document, Applicants authorize the U.S. Patent and Trademark Office to charge any such fees to deposit account No. 50-3355.

Respectfully submitted,

/Caleb Pollack/

Caleb Pollack
Attorney/Agent for Applicant(s)
Registration No. 37,912

Dated: March 29, 2023

Pearl Cohen Zedek Latzer Baratz LLP
7 Times Square, 19th Floor
New York, New York 10036
Tel: (646) 878-0800
Fax: (646) 878-0801

UNITED STATES PATENT AND TRADEMARK OFFICE
COMMISSIONER FOR PATENTS
P.O.BOX 1450
ALEXANDRIA VA 22313-1451

PRESORTED
FIRST-CLASS MAIL
U.S. POSTAGE PAID
POSTEDIGITAL
NNNNN

Pearl Cohen Zedek Latzer Baratz LLP
7 Times Square, 19th Floor
New York, NY 10036



**Courtesy Reminder for
Application Serial No: 16/707,265**

Attorney Docket No: P-587424-US

Customer Number: 49443

Date of Electronic Notification: 02/23/2023

This is a courtesy reminder that new correspondence is available for this application. If you have not done so already, please review the correspondence. The official date of notification of the outgoing correspondence will be indicated on the form PTOL-90 accompanying the correspondence.

An email notification regarding the correspondence was sent to the following email address(es) associated with your customer number:

USPTO@PearlCohen.com

Arch-USPTO@PearlCohen.com

To view your correspondence online or update your email addresses, please visit us anytime at **<https://ppair-my.uspto.gov/pair/PrivatePair>**.

If you have any questions, please email the Electronic Business Center (EBC) at EBC@uspto.gov or call 1-866-217-9197.



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
16/707,265	12/09/2019	Dhruv NAIR	P-587424-US	1015
49443	7590	03/01/2023		
Pearl Cohen Zedek Latzer Baratz LLP			EXAMINER	
7 Times Square, 19th Floor			ALLISON, ANDRAE S	
New York, NY 10036				
			ART UNIT	PAPER NUMBER
			2663	
			NOTIFICATION DATE	DELIVERY MODE
			03/01/2023	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

Arch-USPTO@PearlCohen.com
USPTO@PearlCohen.com

Corrected Notice of Allowability	Application No. 16/707,265	Applicant(s) NAIR et al.	
	Examiner ANDRAE S ALLISON	Art Unit 2663	AIA (FITF) Status Yes

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address--

All claims being allowable, PROSECUTION ON THE MERITS IS (OR REMAINS) CLOSED in this application. If not included herewith (or previously mailed), a Notice of Allowance (PTOL-85) or other appropriate communication will be mailed in due course. **THIS NOTICE OF ALLOWABILITY IS NOT A GRANT OF PATENT RIGHTS.** This application is subject to withdrawal from issue at the initiative of the Office or upon petition by the applicant. See 37 CFR 1.313 and MPEP 1308.

1. ☒ This communication is responsive to 12/09/2019.
☐ A declaration(s)/affidavit(s) under **37 CFR 1.130(b)** was/were filed on _____.
2. ☐ An election was made by the applicant in response to a restriction requirement set forth during the interview on ____; the restriction requirement and election have been incorporated into this action.
3. ☒ The allowed claim(s) is/are See Continuation Sheet. As a result of the allowed claim(s), you may be eligible to benefit from the **Patent Prosecution Highway** program at a participating intellectual property office for the corresponding application. For more information, please see http://www.uspto.gov/patents/init_events/pph/index.jsp or send an inquiry to **PPHfeedback@uspto.gov**.
4. ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

Certified copies:

a) ☐ All b) ☐ Some* c) ☐ None of the:

1. ☐ Certified copies of the priority documents have been received.
2. ☐ Certified copies of the priority documents have been received in Application No. _____.
3. ☐ Copies of the certified copies of the priority documents have been received in this national stage application from the International Bureau (PCT Rule 17.2(a)).

* Certified copies not received: _____.

Applicant has THREE MONTHS FROM THE "MAILING DATE" of this communication to file a reply complying with the requirements noted below. Failure to timely comply will result in ABANDONMENT of this application.
THIS THREE-MONTH PERIOD IS NOT EXTENDABLE.

5. ☐ CORRECTED DRAWINGS (as "replacement sheets") must be submitted.
☐ including changes required by the attached Examiner's Amendment / Comment or in the Office action of Paper No./Mail Date _____.
Identifying indicia such as the application number (see 37 CFR 1.84(c)) should be written on the drawings in the front (not the back) of each sheet. Replacement sheet(s) should be labeled as such in the header according to 37 CFR 1.121(d).
6. ☐ DEPOSIT OF and/or INFORMATION about the deposit of BIOLOGICAL MATERIAL must be submitted. Note the attached Examiner's comment regarding REQUIREMENT FOR THE DEPOSIT OF BIOLOGICAL MATERIAL.

Attachment(s)

1. ☐ Notice of References Cited (PTO-892) 2. ☐ Information Disclosure Statements (PTO/SB/08), Paper No./Mail Date _____. 3. ☐ Examiner's Comment Regarding Requirement for Deposit of Biological Material _____. 4. ☐ Interview Summary (PTO-413), Paper No./Mail Date _____.	5. ☐ Examiner's Amendment/Comment 6. ☐ Examiner's Statement of Reasons for Allowance 7. ☐ Other _____.
---	---

/ANDRAE S ALLISON/ Primary Examiner, Art Unit 2663	
---	--

Continuation of 3. The allowed claim(s) is/are: 7-13 and 18-22 renumbered 1-12

Issue Classification 	Application/Control No. 16/707,265	Applicant(s)/Patent Under Reexamination NAIR et al.
	Examiner ANDRAE S ALLISON	Art Unit 2663

CPC						
Symbol					Type	Version
G06F	/	16	/	2246	F	2019-01-01
G06N	/	3	/	084	I	2013-01-01
G06N	/	20	/	20	I	2019-01-01

CPC Combination Sets							
Symbol				Type	Set	Ranking	Version
	/		/				

NONE		Total Claims Allowed:	
(Assistant Examiner)	(Date)	12	
/ANDRAE S ALLISON/ Primary Examiner, Art Unit 2663	26 February 2023	O.G. Print Claim(s)	O.G. Print Figure
(Primary Examiner)	(Date)	1	1

U.S. Patent and Trademark Office

Part of Paper No.: 20230226

Issue Classification 	Application/Control No. 16/707,265	Applicant(s)/Patent Under Reexamination NAIR et al.
	Examiner ANDRAE S ALLISON	Art Unit 2663

INTERNATIONAL CLASSIFICATION			
CLAIMED			
G06F16/22	/	16	/ 22
G06N3/084	/	3	/ 084
G06N20/20	/	20	/ 20
NON-CLAIMED			
/		/	

US ORIGINAL CLASSIFICATION	
CLASS	SUBCLASS

CROSS REFERENCES(S)						
CLASS	SUBCLASS (ONE SUBCLASS PER BLOCK)					

NONE		Total Claims Allowed:	
(Assistant Examiner)	(Date)	12	
/ANDRAE S ALLISON/ Primary Examiner, Art Unit 2663	26 February 2023	O.G. Print Claim(s)	O.G. Print Figure
(Primary Examiner)	(Date)	1	1

U.S. Patent and Trademark Office

Part of Paper No.: 20230226

Issue Classification 	Application/Control No. 16/707,265	Applicant(s)/Patent Under Reexamination NAIR et al.
	Examiner ANDRAE S ALLISON	Art Unit 2663

☐ Claims renumbered in the same order as presented by applicant ☐ CPA ☐ T.D. ☐ R.1.47															
CLAIMS															
Final	Original	Final	Original	Final	Original	Final	Original	Final	Original	Final	Original	Final	Original	Final	Original
	1	4	10	9	19										
	2	5	11	10	20										
	3	6	12	11	21										
	4	7	13	12	22										
	5		14												
	6		15												
1	7		16												
2	8		17												
3	9	8	18												

NONE (Assistant Examiner) _____ (Date) _____		Total Claims Allowed: 12	
/ANDRAE S ALLISON/ Primary Examiner, Art Unit 2663 (Primary Examiner) _____ (Date) _____		26 February 2023 O.G. Print Claim(s) 1	O.G. Print Figure 1

U.S. Patent and Trademark Office

Part of Paper No.: 20230226



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

NOTICE OF ALLOWANCE AND FEE(S) DUE

49443 7590 02/17/2023
Pearl Cohen Zedek Latzer Baratz LLP
7 Times Square, 19th Floor
New York, NY 10036

EXAMINER

ALLISON, ANDRAE S

ART UNIT

PAPER NUMBER

2663

DATE MAILED: 02/17/2023

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
16/707,265	12/09/2019	Dhruv NAIR	P-587424-US	1015

TITLE OF INVENTION: SYSTEMS AND METHODS FOR PREDICTIVE EARLY STOPPING IN NEURAL NETWORK TRAINING

APPLN. TYPE	ENTITY STATUS	ISSUE FEE DUE	PUBLICATION FEE DUE	PREV. PAID ISSUE FEE	TOTAL FEE(S) DUE	DATE DUE
nonprovisional	SMALL	\$480	\$0.00	\$0.00	\$480	05/17/2023

THE APPLICATION IDENTIFIED ABOVE HAS BEEN EXAMINED AND IS ALLOWED FOR ISSUANCE AS A PATENT. PROSECUTION ON THE MERITS IS CLOSED. THIS NOTICE OF ALLOWANCE IS NOT A GRANT OF PATENT RIGHTS. THIS APPLICATION IS SUBJECT TO WITHDRAWAL FROM ISSUE AT THE INITIATIVE OF THE OFFICE OR UPON PETITION BY THE APPLICANT. SEE 37 CFR 1.313 AND MPEP 1308.

THE ISSUE FEE AND PUBLICATION FEE (IF REQUIRED) MUST BE PAID WITHIN THREE MONTHS FROM THE MAILING DATE OF THIS NOTICE OR THIS APPLICATION SHALL BE REGARDED AS ABANDONED. THIS STATUTORY PERIOD CANNOT BE EXTENDED. SEE 35 U.S.C. 151. THE ISSUE FEE DUE INDICATED ABOVE DOES NOT REFLECT A CREDIT FOR ANY PREVIOUSLY PAID ISSUE FEE IN THIS APPLICATION. IF AN ISSUE FEE HAS PREVIOUSLY BEEN PAID IN THIS APPLICATION (AS SHOWN ABOVE), THE RETURN OF PART B OF THIS FORM WILL BE CONSIDERED A REQUEST TO REAPPLY THE PREVIOUSLY PAID ISSUE FEE TOWARD THE ISSUE FEE NOW DUE.

HOW TO REPLY TO THIS NOTICE:

I. Review the ENTITY STATUS shown above. If the ENTITY STATUS is shown as SMALL or MICRO, verify whether entitlement to that entity status still applies.

If the ENTITY STATUS is the same as shown above, pay the TOTAL FEE(S) DUE shown above.

If the ENTITY STATUS is changed from that shown above, on PART B - FEE(S) TRANSMITTAL, complete section number 5 titled "Change in Entity Status (from status indicated above)".

For purposes of this notice, small entity fees are 40% the amount of undiscounted fees, and micro entity fees are 20% the amount of undiscounted fees.

II. PART B - FEE(S) TRANSMITTAL, or its equivalent, must be completed and returned to the United States Patent and Trademark Office (USPTO) with your ISSUE FEE and PUBLICATION FEE (if required). If you are charging the fee(s) to your deposit account, section "4b" of Part B - Fee(s) Transmittal should be completed. If an equivalent of Part B is filed, a request to reapply a previously paid issue fee must be clearly made, and delays in processing may occur due to the difficulty in recognizing the paper as an equivalent of Part B.

III. All communications regarding this application must give the application number. Please direct all communications prior to issuance to Mail Stop ISSUE FEE unless advised to the contrary.

IMPORTANT REMINDER: Maintenance fees are due in utility patents issuing on applications filed on or after Dec. 12, 1980. It is patentee's responsibility to ensure timely payment of maintenance fees when due. More information is available at www.uspto.gov/PatentMaintenanceFees.

PART B - FEE(S) TRANSMITTAL

Complete and send this form, together with applicable fee(s), by mail or fax, or via EFS-Web.

By mail, send to: Mail Stop ISSUE FEE
Commissioner for Patents
P.O. Box 1450
Alexandria, Virginia 22313-1450

By fax, send to: (571)-273-2885

INSTRUCTIONS: This form should be used for transmitting the ISSUE FEE and PUBLICATION FEE (if required). Blocks 1 through 5 should be completed where appropriate. All further correspondence including the Patent, advance orders and notification of maintenance fees will be mailed to the current correspondence address as indicated unless corrected below or directed otherwise in Block 1, by (a) specifying a new correspondence address; and/or (b) indicating a separate "FEE ADDRESS" for maintenance fee notifications.

CURRENT CORRESPONDENCE ADDRESS (Note: Use Block 1 for any change of address)

49443 7590 02/17/2023
Pearl Cohen Zedek Latzer Baratz LLP
7 Times Square, 19th Floor
New York, NY 10036

Note: A certificate of mailing can only be used for domestic mailings of the Fee(s) Transmittal. This certificate cannot be used for any other accompanying papers. Each additional paper, such as an assignment or formal drawing, must have its own certificate of mailing or transmission.

Certificate of Mailing or Transmission

I hereby certify that this Fee(s) Transmittal is being deposited with the United States Postal Service with sufficient postage for first class mail in an envelope addressed to the Mail Stop ISSUE FEE address above, or being transmitted to the USPTO via EFS-Web or by facsimile to (571) 273-2885, on the date below.

(Typed or printed name)
(Signature)
(Date)

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
16/707,265	12/09/2019	Dhruv NAIR	P-587424-US	1015

TITLE OF INVENTION: SYSTEMS AND METHODS FOR PREDICTIVE EARLY STOPPING IN NEURAL NETWORK TRAINING

APPLN. TYPE	ENTITY STATUS	ISSUE FEE DUE	PUBLICATION FEE DUE	PREV. PAID ISSUE FEE	TOTAL FEE(S) DUE	DATE DUE
nonprovisional	SMALL	\$480	\$0.00	\$0.00	\$480	05/17/2023

EXAMINER	ART UNIT	CLASS-SUBCLASS
ALLISON, ANDRAE S	2663	706-020000

1. Change of correspondence address or indication of "Fee Address" (37 CFR 1.363).

- ☐ Change of correspondence address (or Change of Correspondence Address form PTO/AIA/122 or PTO/SB/122) attached.
- ☐ "Fee Address" indication (or "Fee Address" Indication form PTO/AIA/47 or PTO/SB/47; Rev 03-02 or more recent) attached. **Use of a Customer Number is required.**

2. For printing on the patent front page, list

- (1) The names of up to 3 registered patent attorneys or agents OR, alternatively, 1 _____
- (2) The name of a single firm (having as a member a registered attorney or agent) and the names of up to 2 registered patent attorneys or agents. If no name is listed, no name will be printed. 2 _____
- 3 _____

3. ASSIGNEE NAME AND RESIDENCE DATA TO BE PRINTED ON THE PATENT (print or type)

PLEASE NOTE: Unless an assignee is identified below, no assignee data will appear on the patent. If an assignee is identified below, the document must have been previously recorded, or filed for recordation, as set forth in 37 CFR 3.11 and 37 CFR 3.81(a). Completion of this form is NOT a substitute for filing an assignment.

(A) NAME OF ASSIGNEE

(B) RESIDENCE: (CITY and STATE OR COUNTRY)

Please check the appropriate assignee category or categories (will not be printed on the patent): ☐ Individual ☐ Corporation or other private group entity ☐ Government

4a. Fees submitted: ☐ Issue Fee ☐ Publication Fee (if required) ☐ Advance Order - # of Copies _____

4b. Method of Payment: (Please first reapply any previously paid fee shown above)

- ☐ Electronic Payment via EFS-Web ☐ Enclosed check ☐ Non-electronic payment by credit card (Attach form PTO-2038)
- ☐ The Director is hereby authorized to charge the required fee(s), any deficiency, or credit any overpayment to Deposit Account No. _____

5. Change in Entity Status (from status indicated above)

- ☐ Applicant certifying micro entity status. See 37 CFR 1.29
- ☐ Applicant asserting small entity status. See 37 CFR 1.27
- ☐ Applicant changing to regular undiscounted fee status.

NOTE: Absent a valid certification of Micro Entity Status (see forms PTO/SB/15A and 15B), issue fee payment in the micro entity amount will not be accepted at the risk of application abandonment.

NOTE: If the application was previously under micro entity status, checking this box will be taken to be a notification of loss of entitlement to micro entity status.

NOTE: Checking this box will be taken to be a notification of loss of entitlement to small or micro entity status, as applicable.

NOTE: This form must be signed in accordance with 37 CFR 1.31 and 1.33. See 37 CFR 1.4 for signature requirements and certifications.

Authorized Signature _____

Date _____

Typed or printed name _____

Registration No. _____



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
16/707,265	12/09/2019	Dhruv NAIR	P-587424-US	1015
49443	7590	02/17/2023	EXAMINER	
Pearl Cohen Zedek Latzer Baratz LLP			ALLISON, ANDRAE S	
7 Times Square, 19th Floor			ART UNIT	
New York, NY 10036			PAPER NUMBER	
			2663	
DATE MAILED: 02/17/2023				

Determination of Patent Term Adjustment under 35 U.S.C. 154 (b) (Applications filed on or after May 29, 2000)

The Office has discontinued providing a Patent Term Adjustment (PTA) calculation with the Notice of Allowance.

Section 1(h)(2) of the AIA Technical Corrections Act amended 35 U.S.C. 154(b)(3)(B)(i) to eliminate the requirement that the Office provide a patent term adjustment determination with the notice of allowance. See Revisions to Patent Term Adjustment, 78 Fed. Reg. 19416, 19417 (Apr. 1, 2013). Therefore, the Office is no longer providing an initial patent term adjustment determination with the notice of allowance. The Office will continue to provide a patent term adjustment determination with the Issue Notification Letter that is mailed to applicant approximately three weeks prior to the issue date of the patent, and will include the patent term adjustment on the patent. Any request for reconsideration of the patent term adjustment determination (or reinstatement of patent term adjustment) should follow the process outlined in 37 CFR 1.705.

Any questions regarding the Patent Term Extension or Adjustment determination should be directed to the Office of Patent Legal Administration at (571)-272-7702. Questions relating to issue and publication fee payments should be directed to the Customer Service Center of the Office of Patent Publication at 1-(888)-786-0101 or (571)-272-4200.

OMB Clearance and PRA Burden Statement for PTOL-85 Part B

The Paperwork Reduction Act (PRA) of 1995 requires Federal agencies to obtain Office of Management and Budget approval before requesting most types of information from the public. When OMB approves an agency request to collect information from the public, OMB (i) provides a valid OMB Control Number and expiration date for the agency to display on the instrument that will be used to collect the information and (ii) requires the agency to inform the public about the OMB Control Number's legal significance in accordance with 5 CFR 1320.5(b).

The information collected by PTOL-85 Part B is required by 37 CFR 1.311. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 30 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, Virginia 22313-1450. **DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO:** Commissioner for Patents, P.O. Box 1450, Alexandria, Virginia 22313-1450. Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

Privacy Act Statement

The Privacy Act of 1974 (P.L. 93-579) requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b) (2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C. 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether disclosure of these records is required by the Freedom of Information Act.
2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.
3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.
4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).
5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.
6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).
7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (i.e., GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.
8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspection or an issued patent.
9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

Notice of Allowability	Application No. 16/707,265	Applicant(s) NAIR et al.	
	Examiner ANDRAE S ALLISON	Art Unit 2663	AIA (FITF) Status Yes

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address--

All claims being allowable, PROSECUTION ON THE MERITS IS (OR REMAINS) CLOSED in this application. If not included herewith (or previously mailed), a Notice of Allowance (PTOL-85) or other appropriate communication will be mailed in due course. **THIS NOTICE OF ALLOWABILITY IS NOT A GRANT OF PATENT RIGHTS.** This application is subject to withdrawal from issue at the initiative of the Office or upon petition by the applicant. See 37 CFR 1.313 and MPEP 1308.

1. ☒ This communication is responsive to 12/09/2019.
☐ A declaration(s)/affidavit(s) under **37 CFR 1.130(b)** was/were filed on _____.

2. ☐ An election was made by the applicant in response to a restriction requirement set forth during the interview on _____; the restriction requirement and election have been incorporated into this action.

3. ☒ The allowed claim(s) is/are 1-22. As a result of the allowed claim(s), you may be eligible to benefit from the **Patent Prosecution Highway** program at a participating intellectual property office for the corresponding application. For more information, please see http://www.uspto.gov/patents/init_events/pph/index.jsp or send an inquiry to PPHfeedback@uspto.gov.

4. ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
Certified copies:
a) ☐ All b) ☐ Some* c) ☐ None of the:
1. ☐ Certified copies of the priority documents have been received.
2. ☐ Certified copies of the priority documents have been received in Application No. _____.
3. ☐ Copies of the certified copies of the priority documents have been received in this national stage application from the International Bureau (PCT Rule 17.2(a)).
* Certified copies not received: _____.

Applicant has THREE MONTHS FROM THE "MAILING DATE" of this communication to file a reply complying with the requirements noted below. Failure to timely comply will result in ABANDONMENT of this application.
THIS THREE-MONTH PERIOD IS NOT EXTENDABLE.

5. ☐ CORRECTED DRAWINGS (as "replacement sheets") must be submitted.
☐ including changes required by the attached Examiner's Amendment / Comment or in the Office action of Paper No./Mail Date _____.
Identifying indicia such as the application number (see 37 CFR 1.84(c)) should be written on the drawings in the front (not the back) of each sheet. Replacement sheet(s) should be labeled as such in the header according to 37 CFR 1.121(d).

6. ☐ DEPOSIT OF and/or INFORMATION about the deposit of BIOLOGICAL MATERIAL must be submitted. Note the attached Examiner's comment regarding REQUIREMENT FOR THE DEPOSIT OF BIOLOGICAL MATERIAL.

Attachment(s)

1. ☐ Notice of References Cited (PTO-892)	5. ☐ Examiner's Amendment/Comment
2. ☐ Information Disclosure Statements (PTO/SB/08), Paper No./Mail Date _____.	6. ☒ Examiner's Statement of Reasons for Allowance
3. ☐ Examiner's Comment Regarding Requirement for Deposit of Biological Material _____.	7. ☐ Other _____.
4. ☐ Interview Summary (PTO-413), Paper No./Mail Date. _____.	

/ANDRAE S ALLISON/ Primary Examiner, Art Unit 2663	
---	--

DETAILED ACTION

Notice of Pre-AIA or AIA Status

1. The present application, filed on or after March 16, 2013, is being examined under the first inventor to file provisions of the AIA.

EXAMINER'S AMENDMENT

2. An examiner's amendment to the record appears below. Should the changes and/or additions be unacceptable to applicant, an amendment may be filed as provided by 37 CFR 1.312. To ensure consideration of such an amendment, it **MUST** be submitted no later than the payment of the issue fee.

The application has been amended as follows:

Cancel claims 1-6 and 14-17 (Applicant elect without traverse).

Allowable Subject Matter

3. Claims 7-13 and 18-22 now renumbered 1-12 are allowed.
4. The closest prior art of record is **Prechelt** (NPL titled: Automatic early stopping using cross validation: quantifying the criteria) in view of **Nixon** et al (Pub No.: 2020/0104678). **Prechelt** discloses a method of training a neural network (NN) (automatic early stopping using cross validation – see abstract), the method comprising: over a series of NN training epochs (for e.g. “curve exhibits as many as 16 local minima in the validation set error before severe overfitting begins at about epoch 400” – see section 2, [p][003]), where in **each epoch** the NN undergoes training and a loss is

computed ("For each run, we define $Ev(C)$ as the minimum validation set error found until criterion C indicates to stop; it is the error after epoch number $tm(C)$ (read: 'time of minimum'). $Et(C)$ is the corresponding test set error and characterizes network performance" – see section 6.1, [p][001]): determining, using:

- a set of model parameters (for e.g. network topologies – see section 4, [p][002]);
- data describing training loss of the NN (for e.g. BsubGL2 is normalized test error – see Table 1); and

- a value of improvement in the loss of the NN ("The value $Eopt(t)$ is defined to be the lowest validation set error obtained in epochs up to t : $Eopt(t) = \min_{t' \leq t} Eva(t')$ Now we define the generalization loss at epoch t to be the relative increase of the validation error over the minimum so-far (in percent) - see section 3, [p][003]); and if the probability is less than a threshold, or a wait value is greater than a wait threshold, stopping training ("This leads us to the first class of stopping criteria: stop as soon as the generalization loss exceeds a certain threshold" – see section 3, [p][003]).

However, **Prechelt** does not teach a model that has been trained using training losses of a plurality of NNs other than the NN.

Nixon discloses a method for training optimizer neural network including a model that has been trained using training losses of a plurality of NNs (note that the target network is trained by the optimizer network and a particular neural network – see [p][0050-0051] and [p][0028] [0052] described the loss function associated with the networks).

However, the value disclose by **Prechelt** for determining the loss is an absolute value and not a probability and **Nixon's** optimizer does not exclude the other than the NN; thus the Examiner is withdrawing the rejection and all pending claims are allowed.

The current method improves over the prior art by estimating a predictive early stopping value of a local minima beforehand, and terminate the model training once it is achieved. Example results show embodiments may reduce model training time by an average of 20%, with an average error rate of 4%

5. Any comments considered necessary by applicant must be submitted no later than the payment of the issue fee and, to avoid processing delays, should preferably accompany the issue fee. Such submissions should be clearly labeled "Comments on Statement of Reasons for Allowance."

Inquiries

Any inquiry concerning this communication or earlier communications from the examiner should be directed to ANDRAE S ALLISON whose telephone number is (571)270-1052. The examiner can normally be reached on Monday-Friday 9am-5pm EST.

Examiner interviews are available via telephone, in-person, and video conferencing using a USPTO supplied web-based collaboration tool. To schedule an interview, applicant is encouraged to use the USPTO Automated Interview Request (AIR) at <http://www.uspto.gov/interviewpractice>.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Claire Wang can be reached on (571) 270-1051. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/ANDRAE S ALLISON/

Primary Examiner, Art Unit 2663

February 11, 2023

Issue Classification 	Application/Control No. 16/707,265	Applicant(s)/Patent Under Reexamination NAIR et al.
	Examiner ANDRAE S ALLISON	Art Unit 2663

CPC						
Symbol					Type	Version
G06F	/	16	/	2246	F	2019-01-01
G06N	/	3	/	084	I	2013-01-01
G06N	/	20	/	20	I	2019-01-01

CPC Combination Sets						
Symbol				Type	Set	Ranking
	/		/			

NONE		Total Claims Allowed:	
(Assistant Examiner)	(Date)	12	
/ANDRAE S ALLISON/ Primary Examiner, Art Unit 2663	11 February 2023	O.G. Print Claim(s)	O.G. Print Figure
(Primary Examiner)	(Date)	1	1

U.S. Patent and Trademark Office

Part of Paper No.: 20230211

Issue Classification 	Application/Control No. 16/707,265	Applicant(s)/Patent Under Reexamination NAIR et al.
	Examiner ANDRAE S ALLISON	Art Unit 2663

INTERNATIONAL CLASSIFICATION			
CLAIMED			
G06F16/22	/	16	/ 22
G06N3/084	/	3	/ 084
G06N20/20	/	20	/ 20
NON-CLAIMED			
	/		/

US ORIGINAL CLASSIFICATION	
CLASS	SUBCLASS

CROSS REFERENCES(S)						
CLASS	SUBCLASS (ONE SUBCLASS PER BLOCK)					

NONE		Total Claims Allowed:	
(Assistant Examiner)	(Date)	12	
/ANDRAE S ALLISON/ Primary Examiner, Art Unit 2663	11 February 2023	O.G. Print Claim(s)	O.G. Print Figure
(Primary Examiner)	(Date)	1	1

U.S. Patent and Trademark Office

Part of Paper No.: 20230211

Issue Classification 	Application/Control No. 16/707,265	Applicant(s)/Patent Under Reexamination NAIR et al.
	Examiner ANDRAE S ALLISON	Art Unit 2663

☐ Claims renumbered in the same order as presented by applicant ☐ CPA ☐ T.D. ☐ R.1.47															
CLAIMS															
Final	Original	Final	Original	Final	Original	Final	Original	Final	Original	Final	Original	Final	Original	Final	Original
	1	4	10	9	19										
	2	5	11	10	20										
	3	6	12	11	21										
	4	7	13	12	22										
	5		14												
	6		15												
1	7		16												
2	8		17												
3	9	8	18												

NONE (Assistant Examiner) _____ (Date) _____		Total Claims Allowed: 12	
/ANDRAE S ALLISON/ Primary Examiner, Art Unit 2663 (Primary Examiner) _____ (Date) _____		11 February 2023 O.G. Print Claim(s) 1	O.G. Print Figure 1

U.S. Patent and Trademark Office

Part of Paper No.: 20230211

Search Notes 	Application/Control No. 16/707,265	Applicant(s)/Patent Under Reexamination NAIR et al.
	Examiner ANDRAE S ALLISON	Art Unit 2663

CPC - Searched*		
Symbol	Date	Examiner
G06F16/2246 or G06N3/084 or G06N20/20 or G06N5/003	09/19/2022	AA

CPC Combination Sets - Searched*		
Symbol	Date	Examiner

US Classification - Searched*			
Class	Subclass	Date	Examiner

* See search history printout included with this form or the SEARCH NOTES box below to determine the scope of the search.

Search Notes		
Search Notes	Date	Examiner
Perform inventor search and Perform Assignee search	09/19/2022	AA
NPL Search (keywords are on search reports)	09/19/2022	AA
Consider reference on IDSs	09/19/2022	AA
NPL Search (keywords are on search reports)	02/11/2023	AA
Perform inventor search and Perform Assignee search	02/11/2023	AA

Interference Search			
US Class/CPC Symbol	US Subclass/CPC Group	Date	Examiner
interference	keyword search (see attached search results)	02/11/2023	AA

/ANDRAE S ALLISON/ Primary Examiner, Art Unit 2663	
---	--

Bibliographic Data

Application No: 16/707,265

Foreign Priority claimed: ☐ Yes ☒ No

35 USC 119 (a-d) conditions met: ☐ Yes ☒ No ☐ Met After Allowance

Verified and Acknowledged: /ANDRAE S ALLISON/

Examiner's Signature

Initials

Title:

SYSTEMS AND METHODS FOR PREDICTIVE EARLY
STOPPING IN NEURAL NETWORK TRAINING

FILING or 371(c) DATE	CLASS	GROUP ART UNIT	ATTORNEY DOCKET NO.
12/09/2019	706	2663	P-587424-US
RULE			

APPLICANTS

Comet ML, Inc., New York, NY, UNITED STATES

INVENTORS

Dhruv NAIR, New York, NY, UNITED STATES

Gideon MENDELS, New York, NY, UNITED STATES

Nimrod LAHAV, Hoboken, NJ, UNITED STATES

CONTINUING DATA

This application has PRO of 62852525 05/24/2019

FOREIGN APPLICATIONS

IF REQUIRED, FOREIGN LICENSE GRANTED**

12/30/2019

** SMALL ENTITY **

STATE OR COUNTRY

UNITED STATES

ADDRESS

Pearl Cohen Zedek Latzer Baratz LLP

7 Times Square, 19th Floor

New York, NY 10036

UNITED STATES

FILING FEE RECEIVED

\$1,185

PE2E SEARCH - Search History (Prior Art)

Ref #	Hits	Search Query	DBs	Default Operator	Plurals	British Equivalents	Time Stamp
L37	1	"16707265"	(US-PGPUB; USPAT)	OR	ON	ON	2023/02/11 09:25 PM
L38	142962	(G06F16/2246 or G06N3/084 or G06N20/20 or G06N5/003).cpc.	(US-PGPUB; USPAT; USOCR; FIT (AU, AP, AT, BE, BG, BR, BY, CA, CH, CN, CS, CU, CZ, DD, DE, DK, EA, EE, EP, ES, FI, FR, GB, HR, HU, ID, IE, IL, IS, IT, JP, KR, LT, LU, LV, MA, OA, RU, SU, WO, MC, MD, MY, NL, NO, NZ, PH, PL, PT, RO, RS, SE, SG, SI, SK, TH, TN, TR, TW, UA, VN); FPRS; EPO; JPO; DERWENT; IBM_TDB)	OR	ON	ON	2023/02/11 09:25 PM
L39	7	"10510002"	(US-PGPUB; USPAT)	OR	ON	ON	2023/02/11 09:25 PM
L40	3	("20190050667").pn. OR ("10510002").urpn. AND (PGPB USPT USOC).dbnm.	(US-PGPUB; USPAT; USOCR)	OR	ON	ON	2023/02/11 09:25 PM
L41	402304	(CNN OR (Convolutional NEAR Neural NEAR Network) OR (machine NEAR learning) OR rnn OR (recurrent NEAR neural NEAR network) OR svm OR (support NEAR vector NEAR machine) OR ((deep OR machine) NEAR learning))	(US-PGPUB; USPAT; USOCR)	OR	ON	ON	2023/02/11 09:25 PM
L42	31	5 SAME ((early OR pre\$1mature) NEAR stop\$3)	(US-PGPUB; USPAT; USOCR)	OR	ON	ON	2023/02/11 09:25 PM
L43	672	5 AND ((early OR pre\$1mature) NEAR stop\$3)	(US-PGPUB; USPAT; USOCR)	OR	ON	ON	2023/02/11 09:25 PM
L44	15617	5 AND regularization	(US-PGPUB; USPAT; USOCR)	OR	ON	ON	2023/02/11 09:25 PM
L45	1772	early NEAR stopping	(US-PGPUB; USPAT)	OR	ON	ON	2023/02/11 09:25 PM
L46	292	early NEAR stopping AND 2	(US-PGPUB; USPAT)	OR	ON	ON	2023/02/11 09:25 PM
L47	35	((prevent\$3 OR stop\$3) near3 overfit\$3)	(US-PGPUB; USPAT)	OR	ON	ON	2023/02/11 09:25 PM

02/11/2023 09:29:02 PM
Workspace: 16707265

Page 1 of 4
AA

L48	59	((prevent\$3 OR stop\$3) near3 over\$1fit\$3)	(US-PGPUB; USPAT)	OR	ON	ON	2023/02/11 09:25 PM
L49	3	"11334795"	(US-PGPUB; USPAT)	OR	ON	ON	2023/02/11 09:25 PM
L50	7	("20160224903" "20180357543" "20190095794" "20190236487" "20190392311" "20210174246" "20210190360").pn. OR ("11334795").urpn. AND (PGPB USPT USOC).dbnm.	(US-PGPUB; USPAT; USOCR)	OR	ON	ON	2023/02/11 09:25 PM
L51	3	"11410083"	(US-PGPUB; USPAT; USOCR)	OR	ON	ON	2023/02/11 09:25 PM
L52	14	("5398302" "6269351" "10685286" "20160132787" "20170228645" "20190095794" "20200160118" "20200175354" "20200372342" "20200387798" "20210406695" "20220109853" "20220148245" "20220156518").pn. OR ("11410083").urpn. AND (PGPB USPT USOC).dbnm.	(US-PGPUB; USPAT; USOCR)	OR	ON	ON	2023/02/11 09:25 PM
L53	2	pre\$1train\$3 near3 model SAME early NEAR stop\$3	(US-PGPUB; USPAT)	OR	ON	ON	2023/02/11 09:25 PM
L54	0	base NEAR model near3 model SAME early NEAR stop\$3	(US-PGPUB; USPAT)	OR	ON	ON	2023/02/11 09:25 PM
L55	0	base NEAR model SAME early NEAR stop\$3	(US-PGPUB; USPAT)	OR	ON	ON	2023/02/11 09:25 PM
L56	23	base NEAR model AND early NEAR stop\$3	(US-PGPUB; USPAT)	OR	ON	ON	2023/02/11 09:25 PM
L57	197	expect\$3 near3 train\$3 WITH loss	(US-PGPUB; USPAT)	OR	ON	ON	2023/02/11 09:25 PM
L58	45	expect\$3 near3 train\$3 NEAR loss	(US-PGPUB; USPAT)	OR	ON	ON	2023/02/11 09:25 PM
L59	68	(increas\$3 OR adjust\$3) near3 wait WITH loss	(US-PGPUB; USPAT)	OR	ON	ON	2023/02/11 09:25 PM
L60	0	(increas\$3 OR adjust\$3) near3 patience WITH loss	(US-PGPUB; USPAT)	OR	ON	ON	2023/02/11 09:25 PM

L61	2321	continu\$3 near3 (train\$3 OR learn\$3) WITH (threshold\$3 OR limit OR range)	(US-PGPUB; USPAT)	OR	ON	ON	2023/02/11 09:25 PM
L62	770	continu\$3 near3 (train\$3 OR learn\$3) WITH (threshold\$3 OR limit OR range) SAME (time OR patience)	(US-PGPUB; USPAT)	OR	ON	ON	2023/02/11 09:25 PM
L63	10489	expect\$3 near3 (mean OR averag\$3) AND variance	(US-PGPUB; USPAT)	OR	ON	ON	2023/02/11 09:25 PM
L64	1773	expect\$3 near3 (mean OR averag\$3) SAME variance	(US-PGPUB; USPAT)	OR	ON	ON	2023/02/11 09:25 PM
L65	891	expect\$3 near3 (mean OR averag\$3) WITH variance	(US-PGPUB; USPAT)	OR	ON	ON	2023/02/11 09:25 PM
L66	1	continu\$3 near3 (train\$3 OR learn\$3) WITH patience	(US-PGPUB; USPAT)	OR	ON	ON	2023/02/11 09:25 PM
L67	668	(train\$3 OR learn\$3) WITH patience	(US-PGPUB; USPAT)	OR	ON	ON	2023/02/11 09:25 PM
L68	5	(train\$3 OR learn\$3) WITH patience near3 (limit OR threshold)	(US-PGPUB; USPAT)	OR	ON	ON	2023/02/11 09:25 PM
L69	33	learning NEAR rate WITH patience	(US-PGPUB; USPAT)	OR	ON	ON	2023/02/11 09:25 PM
L70	1214	(increas\$3 OR adjust\$3 OR chang\$3) WITH (patience OR wait\$3) near3 (limit OR threshold)	(US-PGPUB; USPAT)	OR	ON	ON	2023/02/11 09:25 PM
L71	19	(increas\$3 OR adjust\$3 OR chang\$3) WITH (patience OR wait\$3) near3 (limit OR threshold) SAME (train\$3 OR learn\$3)	(US-PGPUB; USPAT)	OR	ON	ON	2023/02/11 09:25 PM
L72	3	"20190354865"	(US-PGPUB; USPAT)	OR	ON	ON	2023/02/11 09:25 PM

PE2E SEARCH - Search History (Interference)

Ref #	Hits	Search Query	DBs	Default Operator	Plurals	British Equivalents	Time Stamp
N1	0	((tran\$3 OR learn\$3) AND (neural NEAR network OR NN OR CNN) AND hyperparameter AND epoch AND (loss OR error) AND probability AND (lower OR below)	(US-PGPUB; USPAT)	OR	ON	ON	2023/02/11 09:28 PM

02/11/2023 09:29:02 PM
Workspace: 16707265

Page 3 of 4
AA

		AND (limit OR threshold OR range) AND wait\$3 AND stop\$3).clm.					
--	--	---	--	--	--	--	--

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Inventor(s):	Dhruv NAIR	Examiner:	ALLISON, Andrae S.
Serial No.:	16/707,265	Group Art Unit:	2663
Filed:	December 9, 2019	Confirmation No.:	1015
Title:	SYSTEMS AND METHODS FOR PREDICTIVE EARLY STOPPING IN NEURAL NETWORK TRAINING		

RESPONSE TO OFFICE ACTION

Commissioner for Patents
P. O. Box 1450
Alexandria, VA 22313-1450

Sir/Madam:

This paper is being filed in response to the Office Action mailed September 27, 2022 by the United States Patent and Trademark Office in connection with the above-identified Application. A response to the September 27, 2022 Office Action was due December 27, 2022. This paper is being filed in conjunction with a petition for a one-month extension of time. Accordingly, this paper is being timely filed.

Remarks/Arguments begin on page **2** of this paper.

APPLICANT(S): Dhruv NAIR
SERIAL NO.: 16/707,265
FILED: December 9, 2019
Page 2

REMARKS

The present response is intended to be fully responsive to all points of objection and/or rejection raised by the Examiner and is believed to place the application in condition for allowance.

Applicants assert that the present invention is new, non-obvious and useful. Prompt consideration and allowance of the claims is respectfully requested.

Status of Claims

Claims 1-22 are pending. Claims 7, 8, 12, 13, 18 and 19 have been rejected. Claims 9-11 and 20-22 have been objected to. Claims 1-6 and 14-17 have been withdrawn.

Restriction Requirement

Applicants elect to prosecute claims 7-13 and 18-22, without traverse. Applicants reserve the right to pursue unelected claims 1-6 and 14-17 in continuation and/or divisional applications.

Claim Objections/Allowable Subject Matter

Claims 9-11 and 20-22 have been objected to. As discussed below, each of Claims 9-11 and 20-22 depends from an allowable independent claim. Thus Applicants request that the objection be withdrawn.

35 U.S.C. § 103 Rejections

In the Office Action, the Examiner rejected claims 7, 13 and 18 under 35 U.S.C. § 103 as being unpatentable over Prechelt ("Automatic Early Stopping Using Cross Validation: Quantifying the Criteria", 1997) in view of Nixon (US Patent Pub. 2020/0104678). The Examiner rejected claims 8 and 19 under 35 U.S.C. § 103 as being unpatentable over Prechelt and Nixon in view of Albon ("Neural Network Entity Early Stopping"). The Examiner rejected claim 12 under 35 U.S.C. § 103 as being unpatentable over Prechelt and Nixon in view of Reisser (US Patent Pub. 2019/0354865).

Applicant respectfully traverses for the reasons that follow.

APPLICANT(S): Dhruv NAIR
SERIAL NO.: 16/707,265
FILED: December 9, 2019
Page 3

Independent claim 7 requires, *inter alia*, “determining ... a probability of improvement in the loss of the NN”. Based in part on this *probability*, claim 7 determines whether or not to stop training: “if the probability is less than a threshold, or a wait value is greater than a wait threshold, stopping training”. For the limitation of a probability the Office Action on page 6 cites to Prechelt, Section 3, p. 3 (it is noted that Section 3 of Prechelt starts on p. 4), specifically Prechelt’s “generalization loss.” Prechelt’s generalization loss is the “relative increase of the validation error over the minimum-so-far”. Prechelt p. 4. Thus Prechelt teaches a generalization loss which is a definite value (not a probability), and which measures increase of error. This does not teach or suggest the claimed *probability* of improvement; rather Prechelt measures an absolute measure of increase.

While each of Prechelt’s measure and the claimed probability may be useful to determine when to stop training, each goes about it in a different manner.

Further, independent claim 7 requires, *inter alia*, that the determination of probability is based on “a model that has been trained using training losses of a plurality of NNs other than the NN” undergoing training and for which a stopping determination is made. For this limitation, on page 7 of the Office Action, the Examiner cites to Nixon, paras. [0050]-[0052] and [0028]. In para. [0028] of Nixon, a task specific loss from the inner network (target NN 120; note it appears a typographic error in para. [0028] refers to target NN 110) is used as a signal to train an outer network (optimizer NN 110). Thus loss from a *single* NN 120 is part of training for NN 110: this does not teach or suggest the claimed use of “a model that has been trained using training losses of a *plurality* of NNs other than the NN” (emphasis added) to determine “a probability that the likely best loss of the NN is lower than the best loss of the other NNs”: in claim 1 a model (not necessarily a NN, see, e.g. para. [0040] of the application as filed) is trained using a plurality of NNs, the model predicting the probability of loss. Nixon paras. [0049]-[0052] trains a NN using data from the same NN in a distributed manner, which also cannot teach or suggest training a model which determines a probability that the likely best loss of the NN is lower than the best loss of the other NNs. Further, Nixon et al. do not use the loss information as input to a model: rather Nixon para. [0028] describes that loss of a NN measures how well NN 120 performs after being

APPLICANT(S): Dhruv NAIR
SERIAL NO.: 16/707,265
FILED: December 9, 2019
Page 4

trained. Thus Nixon's loss is merely signal of a model's performance at predicting parameters for a target network.

None of Albon or Reisser cures the deficiencies of Prechelt and Nixon with respect to claim 7.

While independent claim 18 includes limitations different from those of independent claim 7, the arguments above apply to claim 7 as well.

Each of the dependent claims depends, directly or indirectly, from one of independent claims 7 and 18, and therefore, includes all the limitations of that claim. Accordingly, Applicant respectfully requests the prior art rejections be withdrawn.

Conclusion

In view of the foregoing remarks, Applicants assert that the pending claims are allowable. Their favorable reconsideration and allowance is respectfully requested.

Should the Examiner have any question or comment as to the form, content or entry of this paper, the Examiner is requested to contact the undersigned at the telephone number below. Similarly, if there are any further issues yet to be resolved to advance the prosecution of this application to issue, the Examiner is requested to telephone the undersigned counsel.

To the extent that there are any fees associated with this filing that are due but are not paid or are not able to be paid by a credit card mentioned elsewhere in this document, Applicants authorize the U.S. Patent and Trademark Office to charge any such fees to deposit account No. 50-3355.

Respectfully submitted,

/Caleb Pollack/

Caleb Pollack
Attorney/Agent for Applicant(s)
Registration No. 37,912

APPLICANT(S): Dhruv NAIR
SERIAL NO.: 16/707,265
FILED: December 9, 2019
Page 5

Dated: January 23 2023

Pearl Cohen Zedek Latzer Baratz LLP

1500 Broadway, 12th Floor
New York, New York 10036
Tel: (646) 878-0800
Fax: (646) 878-0801

PETITION FOR EXTENSION OF TIME UNDER 37 CFR 1.136(a)		Docket Number (Optional) P-587424-US																																						
Application Number 16/707,265	Filed December 9, 2019																																							
For SYSTEM AND METHOD FOR EARLY STOPPING WHEN SEARCHING FOR NEURAL NETWORK HYPERPARAMETER DEFINITIONS																																								
Art Unit 2663	Examiner Andrae S. ALLISON																																							
This is a request under the provisions of 37 CFR 1.136(a) to extend the period for filing a reply in the above identified application. The requested extension and fee are as follows (check time period desired and enter the appropriate fee below): <table style="width: 100%; border-collapse: collapse;"><thead><tr><th></th><th style="text-align: center;"><u>Fee</u></th><th style="text-align: center;"><u>Small Entity Fee</u></th><th style="text-align: center;"><u>Micro Entity Fee</u></th><th></th></tr></thead><tbody><tr><td>☒ One month (37 CFR 1.17(a)(1))</td><td style="text-align: center;">\$220</td><td style="text-align: center;">\$88</td><td style="text-align: center;">\$44</td><td style="text-align: center;">\$88</td></tr><tr><td>☐ Two months (37 CFR 1.17(a)(2))</td><td style="text-align: center;">\$640</td><td style="text-align: center;">\$256</td><td style="text-align: center;">\$128</td><td style="text-align: center;">\$</td></tr><tr><td>☐ Three months (37 CFR 1.17(a)(3))</td><td style="text-align: center;">\$1,480</td><td style="text-align: center;">\$592</td><td style="text-align: center;">\$296</td><td style="text-align: center;">\$</td></tr><tr><td>☐ Four months (37 CFR 1.17(a)(4))</td><td style="text-align: center;">\$2,320</td><td style="text-align: center;">\$928</td><td style="text-align: center;">\$464</td><td style="text-align: center;">\$</td></tr><tr><td>☐ Five months (37 CFR 1.17 (a)(5))</td><td style="text-align: center;">\$3,160</td><td style="text-align: center;">\$1,264</td><td style="text-align: center;">\$632</td><td style="text-align: center;">\$</td></tr></tbody></table> ☒ Applicant asserts small entity status. See 37 CFR 1.27. ☐ Applicant certifies micro entity status. See 37 CFR 1.29. Form PTO/SB/15A or B or equivalent must either be enclosed or have been submitted previously. ☐ A check in the amount of the fee is enclosed. ☐ Payment by credit card. Form PTO-2038 is attached. ☐ The Director has already been authorized to charge fees in this application to a Deposit Account. ☒ To the extent that there are any fees associated with this filing that are due but are not paid or are not able to be paid by a credit card mentioned elsewhere in this document, Applicants authorize the U.S. Patent and Trademark Office to charge any such fees to Deposit Account Number 50-3355. ☒ Payment made via EFS-Web. WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038. I am the ☐ applicant / inventor. ☐ assignee of record of the entire interest. See 37 CFR 3.71. 37 CFR 3.73(b) statement is enclosed (Form PTO/SB/96). ☒ attorney or agent of record. Registration number <u>37,912</u> . ☐ attorney or agent under 37 CFR 1.34. Registration number ____. <table style="width: 100%; border-collapse: collapse; margin-top: 20px;"><tr><td style="width: 50%; text-align: center;"><u>/Caleb Pollack/</u></td><td style="width: 50%; text-align: center;"><u>26 January 2023</u></td></tr><tr><td style="text-align: center;">Signature</td><td style="text-align: center;">Date</td></tr><tr><td style="text-align: center;">Caleb Pollack Reg. No. 37,912</td><td style="text-align: center;">(646) 878-0800</td></tr><tr><td style="text-align: center;">Typed or printed name</td><td style="text-align: center;">Telephone Number</td></tr></table> NOTE: This form must be signed in accordance with 37 CFR 1.33. See 37 CFR 1.4 for signature requirements and certifications. Submit multiple forms if more than one signature is required, see below*. ☒ *Total of 1 forms are submitted.				<u>Fee</u>	<u>Small Entity Fee</u>	<u>Micro Entity Fee</u>		☒ One month (37 CFR 1.17(a)(1))	\$220	\$88	\$44	\$88	☐ Two months (37 CFR 1.17(a)(2))	\$640	\$256	\$128	\$	☐ Three months (37 CFR 1.17(a)(3))	\$1,480	\$592	\$296	\$	☐ Four months (37 CFR 1.17(a)(4))	\$2,320	\$928	\$464	\$	☐ Five months (37 CFR 1.17 (a)(5))	\$3,160	\$1,264	\$632	\$	<u>/Caleb Pollack/</u>	<u>26 January 2023</u>	Signature	Date	Caleb Pollack Reg. No. 37,912	(646) 878-0800	Typed or printed name	Telephone Number
	<u>Fee</u>	<u>Small Entity Fee</u>	<u>Micro Entity Fee</u>																																					
☒ One month (37 CFR 1.17(a)(1))	\$220	\$88	\$44	\$88																																				
☐ Two months (37 CFR 1.17(a)(2))	\$640	\$256	\$128	\$																																				
☐ Three months (37 CFR 1.17(a)(3))	\$1,480	\$592	\$296	\$																																				
☐ Four months (37 CFR 1.17(a)(4))	\$2,320	\$928	\$464	\$																																				
☐ Five months (37 CFR 1.17 (a)(5))	\$3,160	\$1,264	\$632	\$																																				
<u>/Caleb Pollack/</u>	<u>26 January 2023</u>																																							
Signature	Date																																							
Caleb Pollack Reg. No. 37,912	(646) 878-0800																																							
Typed or printed name	Telephone Number																																							

This collection of information is required by 37 CFR 1.136(a). The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.11 and 1.14. This collection is estimated to take 6 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

Electronic Patent Application Fee Transmittal				
Application Number:		16707265		
Filing Date:		09-Dec-2019		
Title of Invention:		SYSTEMS AND METHODS FOR PREDICTIVE EARLY STOPPING IN NEURAL NETWORK TRAINING		
First Named Inventor/Applicant Name:		Dhruv NAIR		
Filer:		Caleb Pollack/Lily Li		
Attorney Docket Number:		P-587424-US		
Filed as Small Entity				
Filing Fees for Utility under 35 USC 111(a)				
Description	Fee Code	Quantity	Amount	Sub-Total in USD(\$)
Basic Filing:				
Pages:				
Claims:				
Miscellaneous-Filing:				
Petition:				
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				
Extension-of-Time:				

Description	Fee Code	Quantity	Amount	Sub-Total in USD(\$)
Extension - 1 month with \$0 paid	2251	1	88	88
Miscellaneous:				
Total in USD (\$)				88

Electronic Acknowledgement Receipt	
EFS ID:	47433677
Application Number:	16707265
International Application Number:	
Confirmation Number:	1015
Title of Invention:	SYSTEMS AND METHODS FOR PREDICTIVE EARLY STOPPING IN NEURAL NETWORK TRAINING
First Named Inventor/Applicant Name:	Dhruv NAIR
Customer Number:	49443
Filer:	Caleb Pollack/Lily Li
Filer Authorized By:	Caleb Pollack
Attorney Docket Number:	P-587424-US
Receipt Date:	26-JAN-2023
Filing Date:	09-DEC-2019
Time Stamp:	15:03:19
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	yes
Payment Type	CARD
Payment was successfully received in RAM	\$88
RAM confirmation Number	E20231PF04113377
Deposit Account	503355
Authorized User	Lily Li
The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows: 37 CFR 1.16 (National application filing, search, and examination fees) 37 CFR 1.17 (Patent application and reexamination processing fees)	

37 CFR 1.19 (Document supply fees)					
37 CFR 1.20 (Post Issuance fees)					
37 CFR 1.21 (Miscellaneous fees and charges)					
File Listing:					
Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1		P-587424-US-U-RESOA-EOT-25JAN23.pdf	66243	yes	6
			0ba3c3b03578faa0c44ee7d2aa0d571e3f4f0c		
	Multipart Description/PDF files in .zip description				
	Document Description		Start	End	
	Amendment/Request for Reconsideration-After Non-Final Rejection		1	1	
	Applicant Arguments/Remarks Made in an Amendment		2	5	
	Extension of Time		6	6	
Warnings:					
Information:					
2	Fee Worksheet (SB06)	fee-info.pdf	38503	no	2
			793f2c2d11d4fda8135a9f8164f3c5ef8a95d38e		
Warnings:					
Information:					
Total Files Size (in bytes):			104746		

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
16/707,265	12/09/2019	Dhruv NAIR	P-587424-US	1015
49443	7590	09/27/2022	EXAMINER	
Pearl Cohen Zedek Latzer Baratz LLP			ALLISON, ANDRAE S	
7 Times Square, 19th Floor				
New York, NY 10036			ART UNIT	PAPER NUMBER
			2663	
			NOTIFICATION DATE	DELIVERY MODE
			09/27/2022	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

Arch-USPTO@PearlCohen.com
USPTO@PearlCohen.com

Office Action Summary	Application No. 16/707,265	Applicant(s) NAIR et al.	
	Examiner ANDRAE S ALLISON	Art Unit 2663	AIA (FITF) Status Yes

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTHS FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
 - If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
 - Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133).
 Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1) ☒ Responsive to communication(s) filed on 12/09/2019.
☐ A declaration(s)/affidavit(s) under **37 CFR 1.130(b)** was/were filed on ____.

2a) ☐ This action is **FINAL**. 2b) ☒ This action is non-final.

3) ☐ An election was made by the applicant in response to a restriction requirement set forth during the interview on ____; the restriction requirement and election have been incorporated into this action.

4) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims*

5) ☒ Claim(s) 1-22 is/are pending in the application.
 5a) Of the above claim(s) 1-6 and 14-17 is/are withdrawn from consideration.

6) ☐ Claim(s) ____ is/are allowed.

7) ☒ Claim(s) 7-8, 12-13 and 18-19 is/are rejected.

8) ☒ Claim(s) 9-11 and 20-22 is/are objected to.

9) ☐ Claim(s) ____ are subject to restriction and/or election requirement

* If any claims have been determined allowable, you may be eligible to benefit from the **Patent Prosecution Highway** program at a participating intellectual property office for the corresponding application. For more information, please see http://www.uspto.gov/patents/init_events/pph/index.jsp or send an inquiry to PPHfeedback@uspto.gov.

Application Papers

10) ☐ The specification is objected to by the Examiner.

11) ☒ The drawing(s) filed on 12/09/2019 is/are: a) ☒ accepted or b) ☐ objected to by the Examiner.
 Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
 Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

Priority under 35 U.S.C. § 119

12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
Certified copies:
 a) ☐ All b) ☐ Some** c) ☐ None of the:
 1. ☐ Certified copies of the priority documents have been received.
 2. ☐ Certified copies of the priority documents have been received in Application No. ____.
 3. ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

** See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)

3) ☐ Interview Summary (PTO-413)
 Paper No(s)/Mail Date ____

2) ☒ Information Disclosure Statement(s) (PTO/SB/08a and/or PTO/SB/08b)
 Paper No(s)/Mail Date ____

4) ☐ Other: ____

DETAILED ACTION

Notice of Pre-AIA or AIA Status

1. The present application, filed on or after March 16, 2013, is being examined under the first inventor to file provisions of the AIA.

Information Disclosure Statement

2. The information disclosure statement (IDS) submitted on 12/09/2019 was considered by the examiner. Initialed copy(ies) of the IDS(s) are included.

Election/Restriction

3. Restriction to one of the following inventions is required under 35 U.S.C. 121:
 - I. Claims 1-6 and 14-17, drawn to method for training a neural network including a probability that the likely best loss of the NN is lower than the best loss of the other NN, classified in G06N 20/20.
 - II. Claims 7-13 and 18-22, drawn to method for training a neural network including a probability of improvement in the loss of the NN, classified in G06N 3/084.

The inventions are independent or distinct, each from the other because:

4. Inventions I and II are related as subcombinations disclosed as usable together in a single combination. The subcombinations are distinct if they do not overlap in scope and are not obvious variants, and if it is shown that at least one subcombination is separately usable. In the instant case, subcombination I has separate utility such as a probability that the likely best loss of the NN is lower than the best loss of the other

NN and subcombination II has separate utility such as a probability of improvement in the loss of the NN. See MPEP § 806.05(d).

The examiner has required restriction between subcombinations usable together. Where applicant elects a subcombination and claims thereto are subsequently found allowable, any claim(s) depending from or otherwise requiring all the limitations of the allowable subcombination will be examined for patentability in accordance with 37 CFR 1.104. See MPEP § 821.04(a). Applicant is advised that if any claim presented in a continuation or divisional application is anticipated by, or includes all the limitations of, a claim that is allowable in the present application, such claim may be subject to provisional statutory and/or nonstatutory double patenting rejections over the claims of the instant application.

Restriction for examination purposes as indicated is proper because all the inventions listed in this action are independent or distinct for the reasons given above and there would be a serious search and/or examination burden if restriction were not required because one or more of the following reasons apply:

- (a) the inventions have acquired a separate status in the art in view of their different classification;
- (b) the inventions have acquired a separate status in the art due to their recognized divergent subject matter;
- (c) the inventions require a different field of search (for example, searching different classes/subclasses or electronic resources, or employing different search queries);
- (d) the prior art applicable to one invention would not likely be applicable to

another invention;

Applicant is advised that the reply to this requirement to be complete must include (i) an election of an invention to be examined even though the requirement may be traversed (37 CFR 1.143) **and (ii) identification of the claims encompassing the elected invention.**

The election of an invention may be made with or without traverse. To reserve a right to petition, the election must be made with traverse. If the reply does not distinctly and specifically point out supposed errors in the restriction requirement, the election shall be treated as an election without traverse. Traversal must be presented at the time of election in order to be considered timely. Failure to timely traverse the requirement will result in the loss of right to petition under 37 CFR 1.144. If claims are added after the election, applicant must indicate which of these claims are readable upon the elected invention.

Should applicant traverse on the ground that the inventions are not patentably distinct, applicant should submit evidence or identify such evidence now of record showing the inventions to be obvious variants or clearly admit on the record that this is the case. In either instance, if the examiner finds one of the inventions unpatentable over the prior art, the evidence or admission may be used in a rejection under 35 U.S.C. 103 or pre-AIA 35 U.S.C. 103(a) of the other invention.

During a telephone conversation with Caleb Pollack on September 15, 2022 a provisional election was made without traverse to prosecute the invention of group II, claims 7-13 and 18-22. Affirmation of this election must be made by applicant in replying to this Office action. Claims 1-6 and 14-17 withdrawn from further

consideration by the examiner, 37 CFR 1.142(b), as being drawn to a non-elected invention.

Applicant is reminded that upon the cancellation of claims to a non-elected invention, the inventorship must be corrected in compliance with 37 CFR 1.48(a) if one or more of the currently named inventors is no longer an inventor of at least one claim remaining in the application. A request to correct inventorship under 37 CFR 1.48(a) must be accompanied by an application data sheet in accordance with 37 CFR 1.76 that identifies each inventor by his or her legal name and by the processing fee required under 37 CFR 1.17(i).

Claim Rejections - 35 USC § 103

5. In the event the determination of the status of the application as subject to AIA 35 U.S.C. 102 and 103 (or as subject to pre-AIA 35 U.S.C. 102 and 103) is incorrect, any correction of the statutory basis for the rejection will not be considered a new ground of rejection if the prior art relied upon, and the rationale supporting the rejection, would be the same under either status.

6. The following is a quotation of 35 U.S.C. 103 which forms the basis for all obviousness rejections set forth in this Office action:

A patent for a claimed invention may not be obtained, notwithstanding that the claimed invention is not identically disclosed as set forth in section 102 of this title, if the differences between the claimed invention and the prior art are such that the claimed invention as a whole would have been obvious before the effective filing date of the claimed invention to a person having ordinary skill in the art to which the claimed invention pertains. Patentability shall not be negated by the manner in which the invention was made.

7. Claims 7, 13 and 18 are rejected under 35 U.S.C. 103 as being unpatentable over **Prechelt** (NPL titled: Automatic early stopping using cross validation: quantifying the criteria) in view of **Nixon** et al (Pub No.: 2020/0104678).

As to independent claim 7, **Prechelt** discloses a method of training a neural network (NN) (automatic early stopping using cross validation – see abstract), the method comprising: over a series of NN training epochs (for e.g. “curve exhibits as many as 16 local minima in the validation set error before severe overfitting begins at about epoch 400” – see section 2, [p][003]), where in **each epoch** the NN undergoes training and a loss is computed (“For each run, we define $Ev(C)$ as the minimum validation set error found until criterion C indicates to stop; it is the error after epoch number $tm(C)$ (read: ‘time of minimum’). $Et(C)$ is the corresponding test set error and characterizes network performance” – see section 6.1, [p][001]): determining, using:

a set of model parameters (for e.g. network topologies – see section 4, [p][002]);
data describing training loss of the NN (for e.g. BsubGL2 is normalized test error – see Table 1); and

a probability of improvement in the loss of the NN (“The value $Eopt(t)$ is defined to be the lowest validation set error obtained in epochs up to $t: Eopt(t) = \min_{t' \leq t} Eva(t')$ Now we define the generalization loss at epoch t to be the relative increase of the validation error over the minimum so-far (in percent) - see section 3, [p][003]); and if the probability is less than a threshold, or a wait value is greater than a wait threshold, stopping training (“This leads us to the first class of stopping criteria: stop as soon as the generalization loss exceeds a certain threshold” – see section 3, [p][003]).

However, **Prechelt** does not teach a model that has been trained using training losses of a plurality of NNs other than the NN.

Nixon discloses a method for training optimizer neural network including a model that has been trained using training losses of a plurality of NNs other than the NN (note that the target network is trained by the optimizer network and a particular neural network – see [p][0050-0051] and [p][0028] [0052] described the loss function associated with the networks).

Prechelt & Nixon are combinable because they are directed to machine training.

Before the effective filing date of the claims invention, it would have been obvious to a person of ordinary skill in the art to have considered the method for training optimizer neural network of **Nixon** as a modification to the automatic early stopping using cross validation method of **Prechelt**.

The suggestion/motivation for doing so would have been to automatically determining an operating range of the hyperparameter based on the one or more hyperparameter range trials and training the machine learning model to convergence based at least in part on the determined operating range (see abstract)

Therefore, it would have been obvious to combine **Nixon** with **Prechelt** to obtain the invention as specified in claim 7.

As to claim 13, **Prechelt** teaches the method, wherein model parameters comprise hyperparameters (for e.g. For each run, we define $Ev(C)$ as the minimum validation set error found until criterion C indicates to stop – see section 6.1, [p][001]).

As to claim 18, this claim differs from claim 7, only in that claim 7 is method wherein claim 18 is system and the additional limitations a memory; and a processor are additively recited. **Prechelt** does not expressly disclose a memory; and a processor. **Nixon** discloses a system for training optimizer neural network including a memory (see [p][0074]); and a processor (see [p][0075]). Therefore combining **Prechelt** and **Nixon** would meet the claim limitations for the same reasons as previously discussed in claim 7.

8. Claims 8 and 19 are rejected under 35 U.S.C. 103 as being unpatentable over **Prechelt** (NPL titled: Automatic early stopping using cross validation: quantifying the criteria) in view of **Nixon** et al (Pub No.: 2020/0104678) as applied to claim 7 further in view of **Albon** (NPL titled: Neural Network Early Stopping).

As to claim 8, the combination of **Prechelt** and **Nixon** does not expressly disclose the method, comprising if the probability is not less than a threshold and a wait value is greater than a wait threshold, continuing training.

Albon discloses a neural network early stopping including if the probability is not less than a threshold and a wait value is greater than a wait threshold, continuing training (section Setup Early Stopping - Specifically, in our solution, we included EarlyStopping(monitor='val_loss', patience=2) to define that we wanted to monitor the test (validation) loss at each epoch and after the test loss has not improved after two epochs, training is interrupted. However, since we set patience=2, we won't get the best model, but the model two epochs after the best mode).

Prechelt, Albon & Nixon are combinable because they are directed to machine

training.

Before the effective filing date of the claims invention, it would have been obvious to a person of ordinary skill in the art to have considered the method for t neural network early stopping of **Albon** as a modification to the automatic early stopping using cross validation method of **Prechelt** as modified by **Albon**.

The suggestion/motivation for doing so would have been to implement an early stopping as a callback function at certain stages of the training process, such as at the end of each epoch (section Setup Early Stopping).

Therefore, it would have been obvious to combine **Albon** with **Prechelt** as modified by **Nixon** to obtain the invention as specified in claim 8.

Claim 19 is rejected for the same reasons as set forth in the rejection of the claim 8, as claim 8 is method claim for system claimed in claim 19.

9. Claim 12 is rejected under 35 U.S.C. 103 as being unpatentable over **Prechelt** (NPL titled: Automatic early stopping using cross validation: quantifying the criteria) in view of **Nixon** et al (Pub No.: 2020/0104678) as applied to claim 7 further in view of **REISSER** et al (Pub No.: 2019/0354865).

As to claim 12, the combination of **Prechelt** and **Nixon** as a whole does not wherein determining an expected probability of improvement comprises determining a mean expected training loss and a variance. **REISSER** discloses a learning model including wherein determining an expected probability of improvement comprises determining a mean expected training loss and a variance (see [p][0098] – “the loss

function including a sum of a negative log likelihood function and a cross-entropy function, the negative log likelihood function being a function of the first input and the expected classification, and the cross-entropy function being a function of the expected classification and a mean μ and a variance σ .sup.2 associated with the expected classification”).

Prechelt, REISSER & Nixon are combinable because they are directed to machine training.

Before the effective filing date of the claims invention, it would have been obvious to a person of ordinary skill in the art to have considered the method for learning model of **REISSER** as a modification to the automatic early stopping using cross validation method of **Prechelt** as modified by **Nixon**.

The suggestion/motivation for doing so would have been to determine a stochastic gradient descent repeated until the achievable error rate of the entire system has stopped decreasing or until the error rate has reached a target level (see [p][0043]).

Therefore, it would have been obvious to combine **REISSER** with **Prechelt** as modified by **Nixon** to obtain the invention as specified in claim 12.

Allowable Subject Matter

10. Claims 9-11 and 20-22 are objected to as being dependent upon a rejected base claim, but would be allowable if rewritten in independent form including all of the limitations of the base claim and any intervening claims.

- a. (As to claims 9 and 20) The following is a statement of reasons for the indication of allowable subject matter: None of the prior art of record teaches or

fairly suggest: comprising increasing the wait value if the current loss of the NN is not less than the minimum loss in the loss history for the NN.

b. (As to claims 10 and 21) The following is a statement of reasons for the indication of allowable subject matter: None of the prior art of record teaches or fairly suggest: setting the wait value to zero if the current loss of the NN is less than or equal to the minimum loss in the loss history for the NN.

c. (As to claims 11 and 22) The following is a statement of reasons for the indication of allowable subject matter: None of the prior art of record teaches or fairly suggest: wherein determining using a model a probability of improvement in the loss of the NN comprises obtaining from a leaf of at least one tree data structure data relevant to the NN.

Conclusion

11. The prior art made of record and not relied upon is considered pertinent to applicant's disclosure.

TALATHI et al (Pub No.: 20160224903) discloses HYPER-PARAMETER SELECTION FOR DEEP CONVOLUTIONAL NETWORKS.

. Vasudevan et al (US Patent No.: 11410083) discloses Determining Operating Range Of Hyperparameters.

Inquiries

Any inquiry concerning this communication or earlier communications from the examiner should be directed to ANDRAE S ALLISON whose telephone number is

(571)270-1052. The examiner can normally be reached on Monday-Friday 9am-5pm EST.

Examiner interviews are available via telephone, in-person, and video conferencing using a USPTO supplied web-based collaboration tool. To schedule an interview, applicant is encouraged to use the USPTO Automated Interview Request (AIR) at <http://www.uspto.gov/interviewpractice>.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Claire Wang can be reached on (571) 270-1051. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/ANDRAE S ALLISON/

Primary Examiner, Art Unit 2663

September 19, 2022

<i>Notice of References Cited</i>	Application/Control No. 16/707,265	Applicant(s)/Patent Under Reexamination NAIR et al.	
	Examiner ANDRAE S ALLISON	Art Unit 2663	Page 1 of 1

U.S. PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Name	CPC Classification	US Classification
*	A	US-20200104678-A1	04-2020	Nixon; Jeremy	G06N3/0472	1/1
*	B	US-20190354865-A1	11-2019	REISSER; Matthias	G06N3/0454	1/1
*	C	US-20160224903-A1	08-2016	TALATHI; Sachin Subhash	G06N3/0454	1/1
*	D	US-11410083-B2	08-2022	Vasudevan; Shrihari	G06K9/6289	1/1
*	E	US-20160189210-A1	06-2016	Lacey; Ethan	G06N7/005	705/7.31
*	F	US-20160132787-A1	05-2016	Drevo; Will D.	G06N20/10	706/12
*	G	US-20210174246-A1	06-2021	Triplet; Thomas	G06N3/082	1/1
*	H	US-20190392311-A1	12-2019	Liu; Liu	G06N3/082	1/1
*	I	US-20190236487-A1	08-2019	Huang; Jiapei	G06N20/00	1/1
*	J	US-20200026986-A1	01-2020	HA; SangWon	G06N3/0454	1/1
*	K	US-20200213354-A1	07-2020	CHAKRABORTY; Supriyo	H04L63/0876	1/1
	L					
	M					

FOREIGN PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Country	Name	CPC Classification
	N					
	O					
	P					
	Q					
	R					
	S					
	T					

NON-PATENT DOCUMENTS

*		Include as applicable: Author, Title Date, Publisher, Edition or Volume, Pertinent Pages)
	U	Lutz Prechelt, Automatic early stopping using cross validation: quantifying the criteria, Neural Networks 11 (1998) 761-767 (Year: 1998)
	V	Diaz, Gonzalo I., et al. "An effective algorithm for hyperparameter optimization of neural networks." IBM Journal of Research and Development 61.4/5 (2017): 9-1. (Year: 2017)
	W	Albon, Neural Network Early Stopping, https://chrisalbon.com/code/deep_learning/keras/neural_network_early_stopping/ 12-2017 (Year: 2017)
	X	Jabbar, H., and Rafiqul Zaman Khan. "Methods to avoid over-fitting and under-fitting in supervised machine learning (comparative study)." Computer Science, Communication and Instrumentation Devices 70 (2015). (Year: 2015)

*A copy of this reference is not being furnished with this Office action. (See MPEP § 707.05(a).)
Dates in MM-YYYY format are publication dates. Classifications may be US or foreign.

U.S. Patent and Trademark Office
PTO-892 (Rev. 01-2001)

Notice of References Cited

Part of Paper No. 20220914

<i>Search Notes</i> 	Application/Control No. 16/707,265	Applicant(s)/Patent Under Reexamination NAIR et al.
	Examiner ANDRAE S ALLISON	Art Unit 2663

CPC - Searched*		
Symbol	Date	Examiner
G06F16/2246 or G06N3/084 or G06N20/20 or G06N5/003	09/19/2022	AA

CPC Combination Sets - Searched*		
Symbol	Date	Examiner

US Classification - Searched*			
Class	Subclass	Date	Examiner

* See search history printout included with this form or the SEARCH NOTES box below to determine the scope of the search.

Search Notes		
Search Notes	Date	Examiner
Perform inventor search and Perform Assignee search	09/19/2022	AA
NPL Search (keywords are on search reports)	09/19/2022	AA
Consider reference on IDSs	09/19/2022	AA

Interference Search			
US Class/CPC Symbol	US Subclass/CPC Group	Date	Examiner

/ANDRAE S ALLISON/ Primary Examiner, Art Unit 2663	
---	--

<i>Index of Claims</i> 	Application/Control No. 16/707,265	Applicant(s)/Patent Under Reexamination NAIR et al.
	Examiner ANDRAE S ALLISON	Art Unit 2663

✓	Rejected	-	Cancelled	N	Non-Elected	A	Appeal
=	Allowed	÷	Restricted	I	Interference	O	Objected

CLAIMS										
☐ Claims renumbered in the same order as presented by applicant ☐ CPA ☐ T.D. ☐ R.1.47										
CLAIM			DATE							
Final	Original	09/19/2022								
	1	✓								
	2	✓								
	3	✓								
	4	✓								
	5	✓								
	6	✓								
	7	✓								
	8	✓								
	9	○								
	10	○								
	11	○								
	12	✓								
	13	✓								
	14	✓								
	15	✓								
	16	✓								
	17	✓								
	18	✓								
	19	✓								
	20	○								
	21	○								
	22	○								

PE2E SEARCH - Search History (Prior Art)

Ref #	Hits	Search Query	DBs	Default Operator	Plurals	British Equivalents	Time Stamp
L1	1	"16707265"	(US-PGPUB; USPAT)	OR	ON	ON	2022/08/27 08:46 PM
L2	120676	(G06F16/2246 or G06N3/084 or G06N20/20 or G06N5/003).cpc.	(US-PGPUB; USPAT; USOCR; FIT (AU, AP, AT, BE, BG, BR, BY, CA, CH, CN, CS, CU, CZ, DD, DE, DK, EA, EE, EP, ES, FI, FR, GB, HR, HU, ID, IE, IL, IS, IT, JP, KR, LT, LU, LV, MA, OA, RU, SU, WO, MC, MD, MY, NL, NO, NZ, PH, PL, PT, RO, RS, SE, SG, SI, SK, TH, TN, TR, TW, UA, VN); FPRS; EPO; JPO; DERWENT; IBM_TDB)	OR	ON	ON	2022/08/27 08:47 PM
L3	7	"10510002"	(US-PGPUB; USPAT)	OR	ON	ON	2022/08/27 08:55 PM
L4	3	("20190050667").pn. OR ("10510002").urpn. AND (PGPB USPT USOC).dbnm.	(US-PGPUB; USPAT; USOCR)	OR	ON	ON	2022/08/27 08:55 PM
L5	357732	(CNN OR (Convolutional NEAR Neural NEAR Network) OR (machine NEAR learning) OR rnn OR (recurrent NEAR neural NEAR network) OR svm OR (support NEAR vector NEAR machine) OR ((deep OR machine) NEAR learning))	(US-PGPUB; USPAT; USOCR)	OR	ON	ON	2022/08/27 08:59 PM
L6	25	5 SAME ((early OR pre\$1mature) NEAR stop\$3)	(US-PGPUB; USPAT; USOCR)	OR	ON	ON	2022/08/27 08:59 PM
L7	579	5 AND ((early OR pre\$1mature) NEAR stop\$3)	(US-PGPUB; USPAT; USOCR)	OR	ON	ON	2022/08/27 09:00 PM
L8	13571	5 AND regularization	(US-PGPUB; USPAT; USOCR)	OR	ON	ON	2022/08/27 09:12 PM
L9	1564	early NEAR stopping	(US-PGPUB; USPAT)	OR	ON	ON	2022/09/02 02:10 PM
L10	281	early NEAR stopping AND 2	(US-PGPUB; USPAT)	OR	ON	ON	2022/09/02 02:10 PM
L11	32	((prevent\$3 OR stop\$3) near3 overfit\$3)	(US-PGPUB; USPAT)	OR	ON	ON	2022/09/19 01:07 PM

09/21/2022 01:55:39 PM
Workspace: 16707265

Page 1 of 3
AA

L12	53	((prevent\$3 OR stop\$3) near3 over\$1fit\$3)	(US-PGPUB; USPAT)	OR	ON	ON	2022/09/19 01:10 PM
L13	2	"11334795"	(US-PGPUB; USPAT)	OR	ON	ON	2022/09/19 01:12 PM
L14	7	("20160224903" "20180357543" "20190095794" "20190236487" "20190392311" "20210174246" "20210190360").pn. OR ("11334795").urpn. AND (PGPB USPT USOC).dbnm.	(US-PGPUB; USPAT; USOCR)	OR	ON	ON	2022/09/19 01:14 PM
L15	3	"11410083"	(US-PGPUB; USPAT; USOCR)	OR	ON	ON	2022/09/19 03:11 PM
L16	14	("5398302" "6269351" "10685286" "20160132787" "20170228645" "20190095794" "20200160118" "20200175354" "20200372342" "20200387798" "20210406695" "20220109853" "20220148245" "20220156518").pn. OR ("11410083").urpn. AND (PGPB USPT USOC).dbnm.	(US-PGPUB; USPAT; USOCR)	OR	ON	ON	2022/09/19 03:14 PM
L17	1	pre\$1train\$3 near3 model SAME early NEAR stop\$3	(US-PGPUB; USPAT)	OR	ON	ON	2022/09/19 05:28 PM
L18	0	base NEAR model near3 model SAME early NEAR stop\$3	(US-PGPUB; USPAT)	OR	ON	ON	2022/09/19 05:31 PM
L19	0	base NEAR model SAME early NEAR stop\$3	(US-PGPUB; USPAT)	OR	ON	ON	2022/09/19 05:32 PM
L20	20	base NEAR model AND early NEAR stop\$3	(US-PGPUB; USPAT)	OR	ON	ON	2022/09/19 05:32 PM
L21	170	expect\$3 near3 train\$3 WITH loss	(US-PGPUB; USPAT)	OR	ON	ON	2022/09/20 12:06 PM
L22	42	expect\$3 near3 train\$3 NEAR loss	(US-PGPUB; USPAT)	OR	ON	ON	2022/09/20 12:06 PM
L23	67	(increas\$3 OR adjust\$3) near3 wait WITH loss	(US-PGPUB; USPAT)	OR	ON	ON	2022/09/20 12:14 PM
L24	0	(increas\$3 OR adjust\$3) near3 patience WITH loss	(US-PGPUB; USPAT)	OR	ON	ON	2022/09/20 12:18 PM

L25	2120	continu\$3 near3 (train\$3 OR learn\$3) WITH (threshold\$3 OR limit OR range)	(US-PGPUB; USPAT)	OR	ON	ON	2022/09/20 05:39 PM
L26	706	continu\$3 near3 (train\$3 OR learn\$3) WITH (threshold\$3 OR limit OR range) SAME (time OR patience)	(US-PGPUB; USPAT)	OR	ON	ON	2022/09/20 05:40 PM
L27	10151	expect\$3 near3 (mean OR averag\$3) AND variance	(US-PGPUB; USPAT)	OR	ON	ON	2022/09/20 05:57 PM
L28	1704	expect\$3 near3 (mean OR averag\$3) SAME variance	(US-PGPUB; USPAT)	OR	ON	ON	2022/09/20 05:57 PM
L29	858	expect\$3 near3 (mean OR averag\$3) WITH variance	(US-PGPUB; USPAT)	OR	ON	ON	2022/09/20 05:58 PM
L30	1	continu\$3 near3 (train\$3 OR learn\$3) WITH patience	(US-PGPUB; USPAT)	OR	ON	ON	2022/09/20 06:05 PM
L31	639	(train\$3 OR learn\$3) WITH patience	(US-PGPUB; USPAT)	OR	ON	ON	2022/09/20 06:05 PM
L32	4	(train\$3 OR learn\$3) WITH patience near3 (limit OR threshold)	(US-PGPUB; USPAT)	OR	ON	ON	2022/09/20 06:05 PM
L33	21	learning NEAR rate WITH patience	(US-PGPUB; USPAT)	OR	ON	ON	2022/09/20 06:20 PM
L34	1178	(increas\$3 OR adjust\$3 OR chang\$3) WITH (patience OR wait\$3) near3 (limit OR threshold)	(US-PGPUB; USPAT)	OR	ON	ON	2022/09/21 11:56 AM
L35	15	(increas\$3 OR adjust\$3 OR chang\$3) WITH (patience OR wait\$3) near3 (limit OR threshold) SAME (train\$3 OR learn\$3)	(US-PGPUB; USPAT)	OR	ON	ON	2022/09/21 11:57 AM
L36	1	"20190354865"	(US-PGPUB; USPAT)	OR	ON	ON	2022/09/21 12:18 PM

PE2E SEARCH - Search History (Interference)

There are no Interference searches to show.



Search History

Report Run Date: 2022-09-21 UTC

Searches: 3

Markers: 0

Search #	Timestamp	Results	Page
1	2022-09-19 17:00 UTC	14,732,399	3
2	2022-09-19 17:03 UTC	10,117,176	4
3	2022-09-19 17:11 UTC	10,117,176	4



US Patents | US Designs | US Applications | EPO Patents | EPO Applications | China Patents | China Applications | Japan Patents | Japan Applications | Korea Patents | Korea Applications | WIPO Applications | Argentina Patents | Argentina Applications | Brazil Patents | Brazil Applications | Canada Patents | Canada Applications | Chile Patents | Chile Applications | Colombia Applications | Costa Rica Applications | Cuba Patents | Cuba Applications | Dominican Republic Applications | Ecuador Patents | Ecuador Applications | El Salvador Applications | Guatemala Applications | Honduras Applications | Mexico Patents | Mexico Applications | Nicaragua Patents | Panama Applications | Peru Applications | Trinidad & Tobago Patents | Uruguay Applications | Austria Patents | Austria Applications | Belarus Patents | Belgium Patents | Belgium Applications | Bosnia & Herzegovina Patents | Bosnia & Herzegovina Applications | Bulgaria Patents | Bulgaria Applications | Croatia Patents | Croatia Applications | Czech Republic Patents | Czech Republic Applications | Czechoslovakia Patents | Czechoslovakia Applications | Denmark Patents | Denmark Applications | Estonia Patents | Estonia Applications | EUIPO Patents | Finland Patents | Finland Applications | France Patents | France Applications | Germany Patents | Germany Applications | Great Britain Patents | Great Britain Applications | Greece Patents | Greece Applications | Hungary Patents | Hungary Applications | Iceland Patents | Iceland Applications | Ireland Patents | Ireland Applications | Italy Patents | Italy Applications | Latvia Patents | Latvia Applications | Lithuania Patents | Lithuania Applications | Luxembourg Patents | Malta Patents | Monaco Patents | Montenegro Patents | Montenegro Applications | Netherlands Patents | Netherlands Applications | Norway Patents | Norway Applications | Poland Patents | Poland Applications | Portugal Patents | Portugal Applications | Republic of Moldova Patents | Republic of Moldova Applications | Romania Patents | Romania Applications | San Marino Patents | San Marino Applications | Serbia Patents | Serbia Applications | Slovakia Patents | Slovakia Applications | Slovenia Patents | Spain Patents | Spain Applications | Sweden Patents | Sweden Applications | Switzerland Patents | Switzerland Applications | Ukraine Patents | Yugoslavia/Serbia and Montenegro Patents | Yugoslavia/Serbia and Montenegro Applications | Armenia Patents | Australia Patents | Australia Applications | Cyprus Patents | Gulf Cooperation Council Patents | Hong Kong Patents | India Patents | India Applications | Indonesia Patents | Indonesia Applications | Israel Applications | Jordan Patents | Jordan Applications | Kyrgyzstan Patents | Macao Applications | Malaysia Patents | Mongolia Patents | New Zealand Patents | Philippines Patents | Philippines Applications | Saudi Arabia Patents | Saudi Arabia Applications | Singapore Patents | Singapore Applications | Taiwan Patents | Taiwan Applications | Tajikistan Patents | Tajikistan Applications | Thailand Patents | Thailand Applications | Uzbekistan Patents | Vietnam Patents | Algeria Patents | ARIPO Patents | ARIPO Applications | Egypt Patents | Kenya Patents | Malawi Patents | Morocco Patents | Morocco Applications | OAPI Patents | South Africa Patents | Tunisia Applications | Zambia Patents | Zimbabwe Patents | EAPO Patents | EAPO Applications | Georgia Patents | Georgia Applications | Kazakhstan Patents | Kazakhstan Applications | Russia Patents | Russia Applications | Turkey Patents | Turkey Applications | IEEE Xplore Publications: IEEE Periodicals | IEEE Xplore Publications: IEEE Conferences | IEEE Xplore Publications: IEEE Standards | IEEE Xplore Publications: IEEE Early Access | IEEE Xplore Publications: SMPTE Periodicals | IEEE Xplore Publications: SMPTE Conferences | IEEE Xplore Publications: SMPTE Standards | IEEE Xplore Publications: MIT Press eBooks | IEEE Xplore Publications: Wiley-IEEE eBooks | IEEE Xplore Publications: IBM Periodicals | IEEE Xplore Publications: URSI Periodicals | IEEE Xplore Publications: VDE Conferences | IEEE Xplore Publications: Periodicals from China | IP.com Prior Art Database: The IP.com Journal | IP.com Prior Art Database: Internet Society RFC | IP.com Prior Art Database: IBM TDB Archive | IP.com Prior Art Database: Legacy Journals | IP.com Prior Art Database: Software Patent Institute | OnePetro.org: Periodicals at OnePetro.org | OnePetro.org: Conferences at OnePetro.org | Other Literature: IBM Redbooks | Other Literature: PubMed Central | Other Literature: arXiv.org

1 2022-09-19 17:00 UTC | 14,732,399 results from

De-Dup: None | Cut-off: None | Sort: Relevance

Main Concept: Text
early stopping

InnovationQ | Search History

Page 3 of 4

Concept Modifiers: None

Filters: None

2 2022-09-19 17:03 UTC | 10,117,176 results from 

De-Dup: None | Cut-off: None | Sort: Relevance

Main Concept: Document

 US20200372342 Systems and methods for predictive early stopping in neural network training [\(view\)](#)

Concept Modifiers: None

Filters: None

3 2022-09-19 17:11 UTC | 10,117,176 results from 

De-Dup: None | Cut-off: None | Sort: Relevance

Main Concept: [Same as Search 2](#)

Concept Modifiers (1):

[1. More Like Document:](#)  US20200372342 Systems and methods for predictive early stopping in neural network training [\(view\)](#)

Filters: None

Web Search History

date, time	web site	search string
9/19/2022 10:26:50 AM	Google	marietta ga zip code
9/19/2022 12:00:05 PM	Google	early stopping as nonparametric variational inference
9/19/2022 12:13:26 PM	Google	Early Stopping
9/19/2022 12:13:32 PM	Google	early stopping wait time
9/19/2022 12:13:48 PM	Google	early stopping "wait time"
9/19/2022 12:14:05 PM	Google	deep learning early stopping "wait time"
9/19/2022 12:14:39 PM	Google	deep learning early stopping based on "wait time"
9/19/2022 12:15:08 PM	Google	neural network early stopping
9/19/2022 12:52:51 PM	Google	validation set
9/19/2022 12:53:48 PM	Google	early stopping reference model
9/19/2022 12:55:24 PM	Google	Unsplash
9/19/2022 1:42:21 PM	Google	p.s meaning
9/19/2022 1:52:15 PM	Google	Automatic early stopping using cross validation: quantifying the criteri
9/19/2022 1:52:25 PM	Google	Automatic early stopping using cross validation: quantifying the criteria
9/19/2022 2:05:31 PM	Google	Train faster, generalize better: stability of stochastic gradient descent
9/19/2022 2:21:43 PM	Google	Early stopping — but when?
9/19/2022 3:08:46 PM	Google	Early Stopping with a Validation
9/19/2022 5:29:37 PM	Google	regularization pre\$1trained model
9/19/2022 5:34:23 PM	Google	Keras
9/19/2022 5:35:02 PM	Google	regularization pretrained model
9/19/2022 5:54:15 PM	Google	optimization of neural networks
9/19/2022 5:54:40 PM	Google	optimization of neural networks base model
9/19/2022 5:54:59 PM	Google	regularization of neural networks base model
9/20/2022 12:16:45 PM	Google	increase wait time based on loss
9/20/2022 12:22:53 PM	Google	https://www.google.com/search?q=Understand+the+Impact+of+ Learning +Rate+on+ Neural+ Network+ Performance&rlz=1C1GCEB_enUS780US A8&ved=0ahUKEwitj_u74qP6AhUsFTQIHcvTCf8Q4dUDCA8&uact=5&oq=neural+network+training+increase+wait+time+basec
9/20/2022 6:11:26 PM	Google	learning rate in neural network patience
9/20/2022 6:15:52 PM	Google	https://www.google.com/search?q=neural+network+training+increase+wait+time+based+on+loss&rlz=1C1GCEB_enUS780US A8&ved=0ahUKEwitj_u74qP6AhUsFTQIHcvTCf8Q4dUDCA8&uact=5&oq=neural+network+training+increase+wait+time+basec
9/20/2022 6:26:02 PM	Google	Adamax optimize
9/20/2022 7:58:50 PM	Google	learning rate increase patience
9/20/2022 8:06:20 PM	Google	Stochastic Gradient Descent
9/20/2022 8:06:27 PM	Google	"Stochastic Gradient Descent"
9/21/2022 11:54:14 AM	Google	learning rate "increase patience"
9/21/2022 12:08:40 PM	Google	Neural Network Early Stopping
9/21/2022 12:08:48 PM	Google	"Neural Network Early Stopping"
9/21/2022 12:22:23 PM	Google	learning rate hyperparameter
9/21/2022 1:49:34 PM	Google	An effective algorithm for hyperparameter optimization of neural networks
9/21/2022 1:52:53 PM	Google	METHODS TO AVOID OVERFITTING AND UNDER-FITTING IN SUPERVISED MACHINE LEARNING (COMPARATIVE STUDY

PTO/SB/08b (07-09)
Approved for use through 07/31/2012. OMB 0651-0031
U.S. Patent and Trademark Office: U.S. DEPARTMENT OF COMMERCE

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Substitute for form 1449B/PTO INFORMATION DISCLOSURE STATEMENT BY APPLICANT <i>(use as many sheets as necessary)</i> <i>(use as many sheets as necessary)</i>				Complete if Known	
				Application Number	Not yet assigned
				Filing Date	Herewith
				First Named Inventor	Dhruv NAIR
				Art Unit	Not yet assigned
				Examiner Name	Not yet assigned
Sheet	1	of	1	Attorney Docket Number	P-587424-US

NON PATENT LITERATURE DOCUMENTS			
Examiner Initials*	Cite No. ¹	Include name of the author (in CAPITAL LETTERS), title of the article (where appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published.	T ²
		BAKER et al., "Accelerating Neural Architecture Search Using Performance Prediction", arXiv:1705.10823v2 [cs.LG] November 8, 2017.	☐
		SWERSKY et al., "Freeze-Thaw Bayesian Optimization", arXiv:1406.3896v1 [stat.ML] June 16, 2014.	☐
		KLEIN et al., "Learning Curve Prediction with Bayesian Neural Networks", 2016.	☐
		DOMHAN et al., "Speeding up Automatic Hyperparameter Optimization of Deep Neural Networks by Extrapolation of Learning Curves", Twenty-Fourth International Joint Conference on Artificial Intelligence, 2015.	☐
			☐
			☐
			☐
			☐
			☐
			☐
			☐

Examiner Signature	/ANDRAE S ALLISON/	Date Considered	08/27/2022
-----------------------	--------------------	--------------------	------------

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

¹ Applicant's unique citation designation number (optional). ² Applicant is to place a check mark here if English language Translation is attached. This collection of information is required by 37 CFR 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 2 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 (1-800-786-9199) and select option 2.

Bibliographic Data

Application No: 16/707,265

Foreign Priority claimed: ☐ Yes ☒ No

35 USC 119 (a-d) conditions met: ☐ Yes ☒ No ☐ Met After Allowance

Verified and Acknowledged: /ANDRAE S ALLISON/

Examiner's Signature

Initials

Title:

SYSTEMS AND METHODS FOR PREDICTIVE EARLY
STOPPING IN NEURAL NETWORK TRAINING

FILING or 371(c) DATE	CLASS	GROUP ART UNIT	ATTORNEY DOCKET NO.
12/09/2019	706	2663	P-587424-US
RULE			

APPLICANTS

Comet ML, Inc., New York, NY, UNITED STATES

INVENTORS

Dhruv NAIR, New York, NY, UNITED STATES

Gideon MENDELS, New York, NY, UNITED STATES

Nimrod LAHAV, Hoboken, NJ, UNITED STATES

CONTINUING DATA

This application has PRO of 62852525 05/24/2019

FOREIGN APPLICATIONS

IF REQUIRED, FOREIGN LICENSE GRANTED**

12/30/2019

** SMALL ENTITY **

STATE OR COUNTRY

UNITED STATES

ADDRESS

Pearl Cohen Zedek Latzer Baratz LLP

7 Times Square, 19th Floor

New York, NY 10036

UNITED STATES

FILING FEE RECEIVED

\$1,185



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING OR 371(C) DATE	FIRST NAMED APPLICANT	ATTY. DOCKET NO./TITLE
16/707,265	12/09/2019	Dhruv NAIR	P-587424-US

CONFIRMATION NO. 1015

PUBLICATION NOTICE



OC000000121622350

49443
Pearl Cohen Zedek Latzer Baratz LLP
1500 Broadway
12th Floor
New York, NY 10036

Title:SYSTEMS AND METHODS FOR PREDICTIVE EARLY STOPPING IN NEURAL NETWORK TRAINING

Publication No.US-2020-0372342-A1

Publication Date:11/26/2020

NOTICE OF PUBLICATION OF APPLICATION

The above-identified application will be electronically published as a patent application publication pursuant to 37 CFR 1.211, et seq. The patent application publication number and publication date are set forth above.

The publication may be accessed through the USPTO's publicly available Searchable Databases via the Internet at www.uspto.gov. The direct link to access the publication is currently <http://www.uspto.gov/patft/>.

The publication process established by the Office does not provide for mailing a copy of the publication to applicant. A copy of the publication may be obtained from the Office upon payment of the appropriate fee set forth in 37 CFR 1.19(a)(1). Orders for copies of patent application publications are handled by the USPTO's Public Records Division. The Public Records Division can be reached by telephone at (571) 272-3150 or (800) 972-6382, by facsimile at (571) 273-3250, by mail addressed to the United States Patent and Trademark Office, Public Records Division, Alexandria, VA 22313-1450 or via the Internet.

In addition, information on the status of the application, including the mailing date of Office actions and the dates of receipt of correspondence filed in the Office, may also be accessed via the Internet through the Patent Electronic Business Center at www.uspto.gov using the public side of the Patent Application Information and Retrieval (PAIR) system. The direct link to access this status information is currently <https://portal.uspto.gov/pair/PublicPair>. Prior to publication, such status information is confidential and may only be obtained by applicant using the private side of PAIR.

Further assistance in electronically accessing the publication, or about PAIR, is available by calling the Patent Electronic Business Center at 1-866-217-9197.

Office of Data Management, Application Assistance Unit (571) 272-4000, or (571) 272-4200, or 1-888-786-0101



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
16/707,265	12/09/2019	Dhruv NAIR	P-587424-US	1015
49443	7590	04/07/2020		
Pearl Cohen Zedek Latzer Baratz LLP			EXAMINER	
1500 Broadway			CHAKI, KAKALI	
12th Floor				
New York, NY 10036			ART UNIT	PAPER NUMBER
			2122	
			NOTIFICATION DATE	DELIVERY MODE
			04/07/2020	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

Arch-USPTO@PearlCohen.com
USPTO@PearlCohen.com



UNITED STATES PATENT AND TRADEMARK OFFICE

Commissioner for Patents
United States Patent and Trademark Office
P.O. Box 1450
Alexandria, VA 22313-1450
www.uspto.gov

Pearl Cohen Zedek Latzer Baratz LLP
1500 Broadway
12th Floor
New York, NY 10036

<i>In re</i> Application of: NAIR, Dhruv et al.	:	
Serial No.: 16/707,265	:	
Filed: December 9, 2019	:	
Docket: P-587424-US	:	DECISION ON PETITION UNDER 37
Title: SYSTEMS AND METHODS FOR PREDICTIVE	:	CFR 37 C.F.R §1.184(a)(2) TO
EARLY STOPPING IN NEURAL NETWORK	:	ACCEPT COLOR DRAWINGS
TRAINING	:	
	:	
	:	

This is a decision on the petition filed December 9, 2019 to request acceptance of color drawings under 37 CFR 1.84(a)(2).

The petition is **DISMISSED**.

REGULATION AND PRACTICE

37 C.F.R. 1.84 Standards for drawings.

(a) *Drawings*. There are two acceptable categories for presenting drawings in utility and design patent applications.

(1) *Black ink*. Black and white drawings are normally required. India ink, or its equivalent that secures solid black lines, must be used for drawings; or

(2) *Color*. Color drawings are permitted in design applications. Where a design application contains color drawings, the application must include the number of sets of color drawings required by paragraph (a)(2)(ii) of this section and the specification must contain the reference required by paragraph (a)(2)(iii) of this section. On rare occasions, color drawings may be necessary as the only practical medium by which to disclose the subject matter sought to be patented in a utility patent application. The color drawings must be of sufficient quality such that all details in the drawings are reproducible in black and white in the printed patent. Color drawings are not permitted in international applications (see PCT Rule 11.13). The Office will accept color drawings in utility patent applications only after granting a petition filed under this paragraph explaining why the color drawings are necessary. Any such petition must include the following:

- (i) The fee set forth in § 1.17(h);
- (ii) One (1) set of color drawings if submitted via the Office electronic filing system, or three (3) sets of color drawings if not submitted via the Office electronic filing system; and
- (iii) An amendment to the specification to insert (unless the specification contains or has been previously amended to contain) the following language as the first paragraph of the brief description of the drawings:

The patent or application file contains at least one drawing executed in color. Copies of this patent or patent application publication with color drawing(s) will be provided by the Office upon request and payment of the necessary fee.

DECISION

The petition requests that the color drawings be accepted in the application filed December 9, 2019.

A grantable petition under 37 CFR 1.84(a)(2) must comply with the following requirements:

On rare occasions, color drawings may be necessary as the only practical medium by which to disclose the subject matter sought to be patented in a utility patent application. The Office will accept color drawings in utility patent applications only after granting a petition filed under this paragraph explaining why the color drawings are necessary. Any such petition must include the following:

- (i) The fee set forth in § 1.17(h);
- (ii) One (1) set of color drawings if submitted via the Office electronic filing system, or three (3) sets of color drawings if not submitted via the Office electronic filing system; and
- (iii) An amendment to the specification to insert (unless the specification contains or has been previously amended to contain) the following language as the first paragraph of the brief description of the drawings: "The patent or application file contains at least one drawing executed in color. Copies of this patent or patent application publication with color drawing(s) will be provided by the Office upon request and payment of the necessary fee."

The petition filed on December 9, 2019 to request acceptance of color drawings under 37 CFR 1.84(a)(2) meets requirement (i) as it was filed with the required fee of \$70 on December 9, 2019, (ii) as the one set of color drawings have been filed via EFS-Web on December 9, 2019, and (iii) as the required language is present in the correct location in the specification on page 5, in [0019].

However, a grantable petition must also explain why the color drawing is necessary to depict the subject matter sought to be patented. Petitioner merely states that the color drawings are "necessary as the only practical medium by which aspects of the disclosed subject matter may be accurately conveyed" and "the claimed invention relates to improvements in neural network training, and certain data related to some embodiments can only practically be depicted in the relevant color figures." However, these statements are conclusory and do not explain why the color drawings are necessary to accurately and clearly depict the subject matter sought to be patented.

Petitioner is given one opportunity to perfect the petition. Any request for reconsideration must be filed within TWO MONTHS of the mail date of this decision.

Accordingly, the petition is **DISMISSED**.

Any inquiry regarding this decision should be directed to Brian L. Johnson, Quality Assurance Specialist, at (571) 272-3595.

/Brian L. Johnson/
Brian L. Johnson, Quality Assurance Specialist
Technology Center 2100
Computer Architecture and Software



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING or 371(c) DATE	GRP ART UNIT	FIL FEE REC'D	ATTY DOCKET NO	TOT CLAIMS	IND CLAIMS
16/707,265	12/09/2019	2129	1115	P-587424-US	22	4

CONFIRMATION NO. 1015

FILING RECEIPT



0000000113833395

49443

Pearl Cohen Zedek Latzer Baratz LLP
1500 Broadway
12th Floor
New York, NY 10036

Date Mailed: 01/06/2020

Receipt is acknowledged of this non-provisional utility patent application. The application will be taken up for examination in due course. Applicant will be notified as to the results of the examination. Any correspondence concerning the application must include the following identification information: the U.S. APPLICATION NUMBER, FILING DATE, NAME OF FIRST INVENTOR, and TITLE OF INVENTION. Fees transmitted by check or draft are subject to collection.

Please verify the accuracy of the data presented on this receipt. If an error is noted on this Filing Receipt, please submit a written request for a corrected Filing Receipt, including a properly marked-up ADS showing the changes with strike-through for deletions and underlining for additions. If you received a "Notice to File Missing Parts" or other Notice requiring a response for this application, please submit any request for correction to this Filing Receipt with your reply to the Notice. When the USPTO processes the reply to the Notice, the USPTO will generate another Filing Receipt incorporating the requested corrections provided that the request is grantable.

Inventor(s)

Dhruv NAIR, New York, NY;
Gideon MENDELS, New York, NY;
Nimrod LAHAV, Hoboken, NJ;

Applicant(s)

Comet ML, Inc., New York, NY;

Assignment For Published Patent Application

Comet ML, Inc., New York, NY

Power of Attorney: The patent practitioners associated with Customer Number 49443

Domestic Priority data as claimed by applicant

This appln claims benefit of 62/852,525 05/24/2019

Foreign Applications for which priority is claimed (You may be eligible to benefit from the **Patent Prosecution Highway** program at the USPTO. Please see <http://www.uspto.gov> for more information.) - None.

Foreign application information must be provided in an Application Data Sheet in order to constitute a claim to foreign priority. See 37 CFR 1.55 and 1.76.

Permission to Access Application via Priority Document Exchange: Yes

Permission to Access Search Results: Yes

page 1 of 4

Applicant may provide or rescind an authorization for access using Form PTO/SB/39 or Form PTO/SB/69 as appropriate.

If Required, Foreign Filing License Granted: 12/30/2019

The country code and number of your priority application, to be used for filing abroad under the Paris Convention, is **US 16/707,265**

Projected Publication Date: 11/26/2020

Non-Publication Request: No

Early Publication Request: No

**** SMALL ENTITY ****

Title

SYSTEMS AND METHODS FOR PREDICTIVE EARLY STOPPING IN NEURAL NETWORK
TRAINING

Preliminary Class

706

Statement under 37 CFR 1.55 or 1.78 for AIA (First Inventor to File) Transition Applications: No

PROTECTING YOUR INVENTION OUTSIDE THE UNITED STATES

Since the rights granted by a U.S. patent extend only throughout the territory of the United States and have no effect in a foreign country, an inventor who wishes patent protection in another country must apply for a patent in a specific country or in regional patent offices. Applicants may wish to consider the filing of an international application under the Patent Cooperation Treaty (PCT). An international (PCT) application generally has the same effect as a regular national patent application in each PCT-member country. The PCT process **simplifies** the filing of patent applications on the same invention in member countries, but **does not result** in a grant of "an international patent" and does not eliminate the need of applicants to file additional documents and fees in countries where patent protection is desired.

Almost every country has its own patent law, and a person desiring a patent in a particular country must make an application for patent in that country in accordance with its particular laws. Since the laws of many countries differ in various respects from the patent law of the United States, applicants are advised to seek guidance from specific foreign countries to ensure that patent rights are not lost prematurely.

Applicants also are advised that in the case of inventions made in the United States, the Director of the USPTO must issue a license before applicants can apply for a patent in a foreign country. The filing of a U.S. patent application serves as a request for a foreign filing license. The application's filing receipt contains further information and guidance as to the status of applicant's license for foreign filing.

Applicants may wish to consult the USPTO booklet, "General Information Concerning Patents" (specifically, the section entitled "Treaties and Foreign Patents") for more information on timeframes and deadlines for filing foreign patent applications. The guide is available either by contacting the USPTO Contact Center at 800-786-9199, or it can be viewed on the USPTO website at <http://www.uspto.gov/web/offices/pac/doc/general/index.html>.

For information on preventing theft of your intellectual property (patents, trademarks and copyrights), you may wish to consult the U.S. Government website, <http://www.stopfakes.gov>. Part of a Department of Commerce initiative, this website includes self-help "toolkits" giving innovators guidance on how to protect intellectual property in specific

page 2 of 4

countries such as China, Korea and Mexico. For questions regarding patent enforcement issues, applicants may call the U.S. Government hotline at 1-866-999-HALT (1-866-999-4258).

LICENSE FOR FOREIGN FILING UNDER
Title 35, United States Code, Section 184
Title 37, Code of Federal Regulations, 5.11 & 5.15

GRANTED

The applicant has been granted a license under 35 U.S.C. 184, if the phrase "IF REQUIRED, FOREIGN FILING LICENSE GRANTED" followed by a date appears on this form. Such licenses are issued in all applications where the conditions for issuance of a license have been met, regardless of whether or not a license may be required as set forth in 37 CFR 5.15. The scope and limitations of this license are set forth in 37 CFR 5.15(a) unless an earlier license has been issued under 37 CFR 5.15(b). The license is subject to revocation upon written notification. The date indicated is the effective date of the license, unless an earlier license of similar scope has been granted under 37 CFR 5.13 or 5.14.

This license is to be retained by the licensee and may be used at any time on or after the effective date thereof unless it is revoked. This license is automatically transferred to any related applications(s) filed under 37 CFR 1.53(d). This license is not retroactive.

The grant of a license does not in any way lessen the responsibility of a licensee for the security of the subject matter as imposed by any Government contract or the provisions of existing laws relating to espionage and the national security or the export of technical data. Licensees should apprise themselves of current regulations especially with respect to certain countries, of other agencies, particularly the Office of Defense Trade Controls, Department of State (with respect to Arms, Munitions and Implements of War (22 CFR 121-128)); the Bureau of Industry and Security, Department of Commerce (15 CFR parts 730-774); the Office of Foreign Assets Control, Department of Treasury (31 CFR Parts 500+) and the Department of Energy.

NOT GRANTED

No license under 35 U.S.C. 184 has been granted at this time, if the phrase "IF REQUIRED, FOREIGN FILING LICENSE GRANTED" DOES NOT appear on this form. Applicant may still petition for a license under 37 CFR 5.12, if a license is desired before the expiration of 6 months from the filing date of the application. If 6 months has lapsed from the filing date of this application and the licensee has not received any indication of a secrecy order under 35 U.S.C. 181, the licensee may foreign file the application pursuant to 37 CFR 5.15(b).

SelectUSA

The United States represents the largest, most dynamic marketplace in the world and is an unparalleled location for business investment, innovation, and commercialization of new technologies. The U.S. offers tremendous resources and advantages for those who invest and manufacture goods here. Through SelectUSA, our nation works to promote and facilitate business investment. SelectUSA provides information assistance to the international investor community; serves as an ombudsman for existing and potential investors; advocates on behalf of U.S. cities, states, and regions competing for global investment; and counsels U.S. economic development organizations on investment attraction best practices. To learn more about why the United States is the best country in the world to develop

technology, manufacture products, deliver services, and grow your business, visit <http://www.SelectUSA.gov> or call +1-202-482-6800.

PATENT APPLICATION FEE DETERMINATION RECORD Substitute for Form PTO-875						Application or Docket Number 16/707,265				
APPLICATION AS FILED - PART I										
(Column 1)		(Column 2)		SMALL ENTITY		OR OTHER THAN SMALL ENTITY				
FOR	NUMBER FILED	NUMBER EXTRA	RATE(\$)	FEE(\$)		RATE(\$)	FEE(\$)			
BASIC FEE (37 CFR 1.16(a), (b), or (c))	N/A	N/A	N/A	75		N/A				
SEARCH FEE (37 CFR 1.16(k), (l), or (m))	N/A	N/A	N/A	330		N/A				
EXAMINATION FEE (37 CFR 1.16(o), (p), or (q))	N/A	N/A	N/A	380		N/A				
TOTAL CLAIMS (37 CFR 1.16(i))	22	minus 20 = *	2	x 50 = 100	OR					
INDEPENDENT CLAIMS (37 CFR 1.16(h))	4	minus 3 = *	1	x 230 = 230						
APPLICATION SIZE FEE (37 CFR 1.16(s))	If the specification and drawings exceed 100 sheets of paper, the application size fee due is \$310 (\$155 for small entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).			0.00						
MULTIPLE DEPENDENT CLAIM PRESENT (37 CFR 1.16(j))				0.00						
* If the difference in column 1 is less than zero, enter "0" in column 2.				TOTAL		TOTAL				
				1115						
APPLICATION AS AMENDED - PART II										
(Column 1)		(Column 2)		(Column 3)		SMALL ENTITY		OR OTHER THAN SMALL ENTITY		
AMENDMENT A		CLAIMS REMAINING AFTER AMENDMENT		HIGHEST NUMBER PREVIOUSLY PAID FOR	PRESENT EXTRA	RATE(\$)	ADDITIONAL FEE(\$)		RATE(\$)	ADDITIONAL FEE(\$)
	Total (37 CFR 1.16(i))	*	Minus	**	=	x	=	OR	x	=
	Independent (37 CFR 1.16(h))	*	Minus	***	=	x	=	OR	x	=
	Application Size Fee (37 CFR 1.16(s))							OR		
	FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))							OR		
						TOTAL ADD'L FEE		OR	TOTAL ADD'L FEE	
AMENDMENT B		CLAIMS REMAINING AFTER AMENDMENT		HIGHEST NUMBER PREVIOUSLY PAID FOR	PRESENT EXTRA	RATE(\$)	ADDITIONAL FEE(\$)		RATE(\$)	ADDITIONAL FEE(\$)
	Total (37 CFR 1.16(i))	*	Minus	**	=	x	=	OR	x	=
	Independent (37 CFR 1.16(h))	*	Minus	***	=	x	=	OR	x	=
	Application Size Fee (37 CFR 1.16(s))							OR		
	FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))							OR		
						TOTAL ADD'L FEE		OR	TOTAL ADD'L FEE	
* If the entry in column 1 is less than the entry in column 2, write "0" in column 3. ** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 20, enter "20". *** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 3, enter "3". The "Highest Number Previously Paid For" (Total or Independent) is the highest found in the appropriate box in column 1.										

SCORE Placeholder Sheet for IFW Content

Application Number: 16707265

Document Date: 12/09/2019

The presence of this form in the IFW record indicates that the following document type was received in electronic format on the date identified above. This content is stored in the SCORE database.

Since this was an electronic submission, there is no physical artifact folder, no artifact folder is recorded in PALM, and no paper documents or physical media exist. The TIFF images in the IFW record were created from the original documents that are stored in SCORE.

- Drawing

At the time of document entry (noted above):

- USPTO employees may access SCORE content via DAV or via the SCORE web page.
- External customers may access SCORE content via PAIR using the Supplemental Content tab.

Form Revision Date: March 1, 2019

AUTHORIZATION FOR INTERNET COMMUNICATIONS IN A PATENT APPLICATION OR REQUEST TO WITHDRAW AUTHORIZATION FOR INTERNET COMMUNICATIONS	Application No.	Not yet assigned
	Filing Date	Herewith
	First Named Inventor	Dhruv NAIR
	Art Unit	Not yet assigned
	Examiner Name	Not yet assigned
	Practitioner Docket No.	P-587424-US

To: Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

I. To authorize permission for Internet Communications.

☒ Recognizing that Internet communications are not secure, I hereby authorize the USPTO to communicate with the undersigned and practitioners in accordance with 37 CFR 1.33 and 37 CFR 1.34 concerning any subject matter of this application via video conferencing, instant messaging, or electronic mail. I understand that a copy of these communications will be made of record in the application file. (MPEP 502.03)

II. To withdraw authorization for Internet Communications.

☐ The authorization given on _____, to the USPTO to communicate with the undersigned and any practitioner in accordance with 37 CFR 1.33 and 37 CFR 1.34 concerning any subject matter of this application via Internet communications such as video conferencing, instant messaging, or electronic mail is hereby withdrawn. I understand that the withdrawal is effective when approved rather than when received.

I am the

☐ applicant.

☒ attorney or agent of record. Registration number 37,912

☐ attorney or agent acting under 37 CFR 1.34. Registration number _____

/Caleb Pollack/
Signature

9 December 2019
Date

Caleb Pollack Reg. No. 37,912
Typed or printed name

(646) 878-0800
Telephone Number

NOTE: This form must be signed in accordance with 37 CFR 1.33. See 37 CFR 1.4 for signature requirements and certifications. Juristic entities must be represented by a patent practitioner (see 37 CFR 1.31, which is applicable to any paper filed on or after September 16, 2012, that is presented on behalf of a juristic entity, regardless of application filing date). Submit multiple forms if more than one signature is required, see below*.

☒ *Total of 1 forms are submitted.

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Inventors:	Dhruv NAIR et al.	Examiner:	Not yet assigned
Serial No.:	Not yet assigned	Group Art Unit:	Not yet assigned
Filed:	Herewith	Confirmation No.:	Not yet assigned
Title:	SYSTEMS AND METHODS FOR PREDICTIVE EARLY STOPPING IN NEURAL NETWORK TRAINING		

Commissioner for Patents
P. O. Box 1450
Alexandria, VA 22313-1450

PETITION UNDER 37 CFR 1.84(a)(2) TO ACCEPT COLOR DRAWINGS

Sir/Madam:

Color drawings have been submitted in this application. The color drawings are necessary as the only practical medium by which aspects of the disclosed subject matter may be accurately conveyed. For example, the claimed invention relates to improvements in neural network training, and certain data related to some embodiments can only practically be depicted in the relevant color figures.

The first paragraph of the Brief Description of the Drawings section of the Specification has been amended to include the appropriate language.

As the color drawings are being filed electronically via EFS-Web, only one set of the drawings is submitted.

The requisite fee for this Petition is to be paid by credit card (deficiencies to be paid by the below authorization to charge Deposit Account **50-3355**).

To the extent that there are any fees associated with this filing that are due but are not paid or are not able to be paid by a credit card mentioned elsewhere in this document, Applicants

authorize the U.S. Patent and Trademark Office to charge any such fees to deposit account No. **50-3355**.

Respectfully submitted,

/Caleb Pollack/

Caleb Pollack

Attorney/Agent for Applicant(s)

Registration No. 37,912

Dated: December 9, 2019

Pearl Cohen Zedek Latzer Baratz LLP

1500 Broadway, 12th Floor

New York, New York 10036

Tel: (646) 878-0800

Fax: (646) 878-0801

STATEMENT UNDER 37 CFR 3.73(c)

Applicant/Patent Owner: **Dhruv NAIR et al.** Docket No.: **P-587424-US**

Application No./Patent No.: **Not yet assigned** Filed/Issue Date: **Herewith**

Titled: **SYSTEMS AND METHODS FOR PREDICTIVE EARLY STOPPING IN NEURAL NETWORK TRAINING**

Comet ML, Inc., a **company**,

(Name of Assignee) (Type of Assignee, e.g., corporation, partnership, university, government agency, etc.)

states that, for the patent application/patent identified above, it is (choose **one** of options 1, 2, 3 or 4 below):

1. ☒ The assignee of the entire right, title, and interest.
2. ☐ An assignee of less than the entire right, title, and interest (check applicable box):
 - ☐ The extent (by percentage) of its ownership interest is ____ %. Additional Statement(s) by the owners holding the balance of the interest must be submitted to account for 100% of the ownership interest.
 - ☐ There are unspecified percentages of ownership. The other parties, including inventors, who together own the entire right, title and interest are:

Additional Statement(s) by the owner(s) holding the balance of the interest must be submitted to account for the entire right, title, and interest.
3. ☐ The assignee of an undivided interest in the entirety (a complete assignment from one of the joint inventors was made). The other parties, including inventors, who together own the entire right, title, and interest are:

Additional Statement(s) by the owner(s) holding the balance of the interest must be submitted to account for the entire right, title, and interest.
4. ☐ The recipient, via a court proceeding or the like (e.g., bankruptcy, probate), of an undivided interest in the entirety (a complete transfer of ownership interest was made). The certified document(s) showing the transfer is attached.

The interest identified in option 1, 2 or 3 above (not option 4) is evidenced by either (choose **one** of options A or B below):

- A. ☒ An assignment from the inventor(s) of the patent application/patent identified above. The assignment was recorded in the United States Patent and Trademark Office at Reel ____, Frame ____, or for which a copy thereof is attached.
- B. ☐ A chain of title from the inventor(s), of the patent application/patent identified above, to the current assignee as follows:
 1. From: _____ To: _____

The document was recorded in the United States Patent and Trademark Office at
Reel ____, Frame ____, or for which a copy thereof is attached.
 2. From: _____ To: _____

The document was recorded in the United States Patent and Trademark Office at
Reel ____, Frame ____, or for which a copy thereof is attached.

[Page 1 of 2]

This collection of information is required by 37 CFR 3.73(b). The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.11 and 1.14. This collection is estimated to take 12 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

STATEMENT UNDER 37 CFR 3.73(c)

3. From: _____ To: _____

The document was recorded in the United States Patent and Trademark Office at

Reel ____, Frame ____, or for which a copy thereof is attached.

4. From: _____ To: _____

The document was recorded in the United States Patent and Trademark Office at

Reel ____, Frame ____, or for which a copy thereof is attached.

5. From: _____ To: _____

The document was recorded in the United States Patent and Trademark Office at

Reel ____, Frame ____, or for which a copy thereof is attached.

6. From: _____ To: _____

The document was recorded in the United States Patent and Trademark Office at

Reel ____, Frame ____, or for which a copy thereof is attached.

☐ Additional documents in the chain of title are listed on a supplemental sheet(s).☐ As required by 37 CFR 3.73(c)(1)(i), the documentary evidence of the chain of title from the original owner to the assignee was, or concurrently is being, submitted for recordation pursuant to 37 CFR 3.11.

[NOTE: A separate copy (i.e., a true copy of the original assignment document(s)) must be submitted to Assignment Division in accordance with 37 CFR Part 3, to record the assignment in the records of the USPTO. See MPEP 302.08]

The undersigned (whose title is supplied below) is authorized to act on behalf of the assignee.

/Caleb Pollack/
Signature9 December 2019
DateCaleb Pollack
Printed or Typed Name37,912
Title or Registration Number

[Page 2 of 2]

Docket No.: P-587424-US**ASSIGNMENT**

In consideration of One Dollar (\$1.00), and other good and valuable consideration, the Receipt of which is hereby acknowledged, I the undersigned, Dhruv NAIR residing at 240 Richardson St, Apt 3A, New York, NY 11222, USA; Gideon MENDELS residing at 1 University Place 2B, New York, NY 10003, USA; Nimrod LAHAV residing at 333 River Street, Apt 508, Hoboken, NJ 07030, USA (hereinafter, the "Assignor")

Hereby sell, assign and transfer to Comet ML, Inc., a company located at 100 Avenue of the Americas, 3rd Floor, New York, NY 10013, USA its successors, assigns and legal representatives (hereinafter, the "Assignee"), the entire right, title and interest for all countries, in and to any and all inventions which are disclosed and claimed, and as possessed by the undersigned, any and all inventions which are disclosed but not claimed in United States Patent Application No. (_____) filed on (_____) and being entitled:

**SYSTEMS AND METHODS FOR PREDICTIVE EARLY STOPPING IN NEURAL
NETWORK TRAINING**

and in and to said application and all divisional, continuing, substitute, renewal, reissue, and all other applications for U.S. Letters Patent or other related property rights in any and all foreign countries which have been or shall be filed on any of said inventions disclosed in said application (including but not limited to US provisional patent application 62/852,525 filed on May 24, 2019); and in and to all original and reissued patents or related foreign documents which have been or shall be issued on said inventions; and including the Assignor's right in and to all revenue, income, royalties, damages and payments now or hereafter due or payable with respect to any Letters Patent which may be granted, and in and to all causes of action (either in law or in equity), and the right to sue, counterclaim, and recover for past, present and future infringement of the rights assigned or to be assigned under this Assignment. I hereby authorize and request the attorney of record of the firm Pearl Cohen Zedek Latzer Baratz LLP, 1500 Broadway, 12th Floor, New York, New York 10036, to insert in the parentheses above, the filing date and US Serial No. of said application when known;

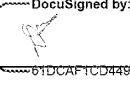
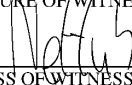
Authorize and request the Commissioner of Patents of the United States to issue to said Assignee, the corporation above named, its successors, assigns and legal representatives, in accordance with this assignment, any and all United States Letters Patent on said inventions or any of them disclosed in said application;

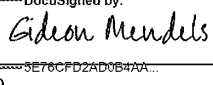
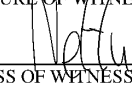
Agree that said Assignee may apply for and receive foreign Letters Patent or rights of any other kind for said inventions, or any of them; and may claim, in applications for said foreign Letters Patent

Docket No.: P-587424-US

or other rights, the priority of the aforesaid United States patent application under the provisions of the International Convention of 1883 and later modifications thereof, under the Patent Cooperation Treaty, under the European Patent Convention or under any other available international agreement; and that, when requested, without charge to, but at the expense of, said Assignee, its successors, assigns and legal representatives, to carry out in good faith the intent and purpose of this assignment, the under-signed or the undersigned's executors or administrators will, for the United States and all foreign countries, execute all divisional, continuing, substitute, renewal, reissue, and all other patent applications or other documents on any and all said inventions; execute all rightful oaths, assignments, powers of attorney and other papers; communicate to said Assignee, its successors, assigns and representatives, all facts known and documents available to the undersigned relating to said inventions and the history thereof; testify in all legal proceedings; and generally do everything possible which said Assignee, its successors, assigns or representatives shall consider desirable for aiding in securing, maintaining and enforcing proper patent protection for said inventions and for vesting title to said inventions and all applications for patents or related foreign rights and all patents on said inventions, in said Assignee, its successors, assigns and legal representatives; and

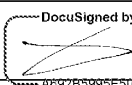
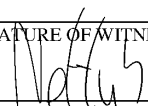
COVENANT with said Assignee, its successors, assigns and legal representatives that no assignment, grant, mortgage, license or other agreement affecting the rights and property herein conveyed has been made to others by the undersigned, and that full right to convey the same as herein expressed is possessed by the undersigned.

FULL NAME OF INVENTOR: Dhruv NAIR	FULL NAME OF WITNESS: Nettah Lee Horesh
SIGNATURE OF INVENTOR:  DocuSigned by: 61DCAF1CD44941F...	SIGNATURE OF WITNESS: 
DATE: <u>05/12/2019</u> (day / month / year)	ADDRESS OF WITNESS: 603 Willow Ave, Hoboken, 07030 NJ

FULL NAME OF INVENTOR: Gideon MENDELS	FULL NAME OF WITNESS: Nettah Lee Horesh
SIGNATURE OF INVENTOR:  DocuSigned by: 5E76CFD2AD0B4AA...	SIGNATURE OF WITNESS: 
DATE: <u>05/12/2019</u> (day / month / year)	ADDRESS OF WITNESS: 603 Willow Ave, Hoboken, 07030 NJ

- 2 -

Docket No.: P-587424-US

FULL NAME OF INVENTOR: Nimrod LAHAV	FULL NAME OF WITNESS: Nettah Lee Horesh
SIGNATURE OF INVENTOR:  A692B5995E5049D...	SIGNATURE OF WITNESS: 
DATE: <u>05/12/2019</u> (day / month / year)	ADDRESS OF WITNESS: 603 Willow Ave, Hoboken, 07030 NJ

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

POWER OF ATTORNEY TO PROSECUTE APPLICATIONS BEFORE THE USPTO

I hereby revoke all previous powers of attorney given in the application identified in the attached statement under 37 CFR 3.73(c).

I hereby appoint:

☒ Practitioners associated with the Customer Number:

49443

OR☐ Practitioner(s) named below (if more than ten patent practitioners are to be named, then a customer number must be used):

Name Registration	Number		Name Registration	Number

as attorney(s) or agent(s) to represent the undersigned before the United States Patent and Trademark Office (USPTO) in connection with any and all patent applications assigned only to the undersigned according to the USPTO assignment records or assignment documents attached to this form in accordance with 37 CFR 3.73(c).

Please change the correspondence address for the application identified in the attached statement under 37 CFR 3.73(c) to:

☒ The address associated with Customer Number:

49443

OR☐ Firm or
Individual Name

Pearl Cohen Zedek Latzer Baratz LLP

Address 1500 Broadway, 12th Floor

City New York

State NY

ZIP 10036

Country USA

Telephone (646) 878-0800

Email USPTO@PearlCohen.com

Assignee Name and Address:

Comet ML, Inc.
100 Avenue of the Americas, 3rd Floor
New York
NY 10013
USA**A copy of this form, together with a statement under 37 CFR 3.73(c) (Form PTO/SB/96 or equivalent) is required to be filed in each application in which this form is used. The statement under 37 CFR 3.73(c) may be completed by one of the practitioners appointed in this form, and must identify the application in which this Power of Attorney is to be filed.****SIGNATURE of Assignee of Record**

The individual whose signature and title is supplied below is authorized to act on behalf of the assignee

Signature		Date	12/3/19
Name	Gideon Mendels	Telephone	347-247-0493
Title	CEO		

This collection of information is required by 37 CFR 1.31, 1.32 and 1.33. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.11 and 1.14. This collection is estimated to take 3 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

**DECLARATION (37 CFR 1.63) FOR UTILITY OR DESIGN APPLICATION USING AN
APPLICATION DATA SHEET (37 CFR 1.76)**

Title of Invention	Attorney Docket No: P-587424-US SYSTEMS AND METHODS FOR PREDICTIVE EARLY STOPING IN NEURAL NETWORK TRAINING
As the below named inventor, I hereby declare that: This declaration is directed to: ☒ The attached application, or ☐ United States application or PCT international application number filed on _. The above-identified application was made or authorized to be made by me. I believe that I am the original inventor or an original joint inventor of a claimed invention in the application I hereby acknowledge that any willful false statement made in this declaration is punishable under 18 U.S.C. 1001 by fine or imprisonment of not more than five (5) years, or both. WARNING: Petitioner/applicant is cautioned to avoid submitting personal information in documents filed in a patent application that may contribute to identity theft. Personal information such as social security numbers, bank account numbers, or credit card numbers (other than a check or credit card authorization form PTO-2038 submitted for payment purposes) is never required by the USPTO to support a petition or an application. If this type of personal information is included in documents submitted to the USPTO, petitioners/applicants should consider redacting such personal information from the documents before submitting them to the USPTO. Petitioner/applicant is advised that the record of a patent application is available to the public after publication of the Application (unless a non-publication request in compliance with 37 CFR 1.213(a) is made in the application) or issuance of a Patent. Furthermore, the record from an abandoned application may also be available to the public if the application is Referenced in a published application or an issued patent (see 37 CFR 1.14). Checks and credit card authorization forms PTO-2038 submitted for payment purposes are not retained in the application file and therefore are not publicly available. LEGAL NAME OF INVENTOR Inventor: <u>Dhruv NAIR</u> (Date Optional): <u>12/5/2019</u> Signature:  DocuSigned by: <u>61DCAF1CD44941F...</u> Note: An application data sheet (PTO/SB/14 or equivalent), including naming the entire inventive entity, must accompany this form or must have been previously filed. Use an additional PTO/AIA/01 form for each additional inventor.	

This collection of information is required by 35 U.S.C. 115 and 37 CFR 1.63. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.11 and 1.14. This collection is estimated to take 1 minute to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

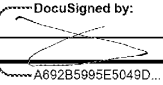
Title of Invention	SYSTEMS AND METHODS FOR PREDICTIVE EARLY STOPING IN NEURAL NETWORK TRAINING
As the below named inventor, I hereby declare that: This declaration is directed to: ☒ ☐ The attached application, or United States application or PCT international application number filed on ___. The above-identified application was made or authorized to be made by me. I believe that I am the original inventor or an original joint inventor of a claimed invention in the application I hereby acknowledge that any willful false statement made in this declaration is punishable under 18 U.S.C. 1001 by fine or imprisonment of not more than five (5) years, or both. WARNING: Petitioner/applicant is cautioned to avoid submitting personal information in documents filed in a patent application that may contribute to identity theft. Personal information such as social security numbers, bank account numbers, or credit card numbers (other than a check or credit card authorization form PTO-2038 submitted for payment purposes) is never required by the USPTO to support a petition or an application. If this type of personal information is included in documents submitted to the USPTO, petitioners/applicants should consider redacting such personal information from the documents before submitting them to the USPTO. Petitioner/applicant is advised that the record of a patent application is available to the public after publication of the Application (unless a non-publication request in compliance with 37 CFR 1.213(a) is made in the application) or issuance of a Patent. Furthermore, the record from an abandoned application may also be available to the public if the application is Referenced in a published application or an issued patent (see 37 CFR 1.14). Checks and credit card authorization forms PTO-2038 submitted for payment purposes are not retained in the application file and therefore are not publicly available.	
LEGAL NAME OF INVENTOR Inventor: <u>Gideon MENDELS</u> (Date Optional): <u>12/5/2019</u> Signature: <u></u> DocuSigned by: 5E76CFD2AD0B4AA...	
Note: An application data sheet (PTO/SB/14 or equivalent), including naming the entire inventive entity, must accompany this form or must have been previously filed. Use an additional PTO/AIA/01 form for each additional inventor.	

This collection of information is required by 35 U.S.C. 115 and 37 CFR 1.63. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.11 and 1.14. This collection is estimated to take 1 minute to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

**DECLARATION (37 CFR 1.63) FOR UTILITY OR DESIGN APPLICATION USING AN
APPLICATION DATA SHEET (37 CFR 1.76)**

Title of Invention	Attorney Docket No: P-587424-US
SYSTEMS AND METHODS FOR PREDICTIVE EARLY STOPING IN NEURAL NETWORK TRAINING	
As the below named inventor, I hereby declare that: This declaration is directed to: ☒ The attached application, or ☐ United States application or PCT international application number filed on . The above-identified application was made or authorized to be made by me. I believe that I am the original inventor or an original joint inventor of a claimed invention in the application I hereby acknowledge that any willful false statement made in this declaration is punishable under 18 U.S.C. 1001 by fine or imprisonment of not more than five (5) years, or both. WARNING: Petitioner/applicant is cautioned to avoid submitting personal information in documents filed in a patent application that may contribute to identity theft. Personal information such as social security numbers, bank account numbers, or credit card numbers (other than a check or credit card authorization form PTO-2038 submitted for payment purposes) is never required by the USPTO to support a petition or an application. If this type of personal information is included in documents submitted to the USPTO, petitioners/applicants should consider redacting such personal information from the documents before submitting them to the USPTO. Petitioner/applicant is advised that the record of a patent application is available to the public after publication of the Application (unless a non-publication request in compliance with 37 CFR 1.213(a) is made in the application) or issuance of a Patent. Furthermore, the record from an abandoned application may also be available to the public if the application is Referenced in a published application or an issued patent (see 37 CFR 1.14). Checks and credit card authorization forms PTO-2038 submitted for payment purposes are not retained in the application file and therefore are not publicly available.	
LEGAL NAME OF INVENTOR Inventor: <u>Nimrod LAHAV</u> (Date Optional): <u>12/5/2019</u> Signature: _____ 	
Note: An application data sheet (PTO/SB/14 or equivalent), including naming the entire inventive entity, must accompany this form or must have been previously filed. Use an additional PTO/AIA/01 form for each additional inventor.	

This collection of information is required by 35 U.S.C. 115 and 37 CFR 1.63. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.11 and 1.14. This collection is estimated to take 1 minute to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

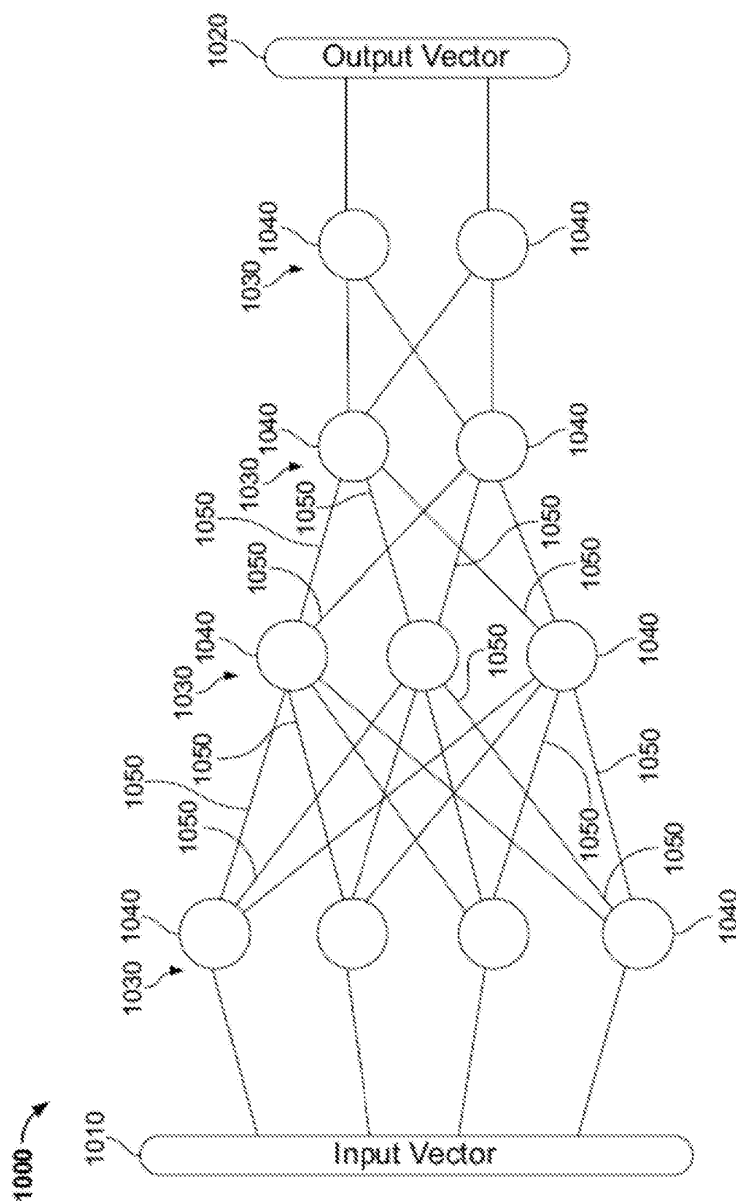


Fig. 1

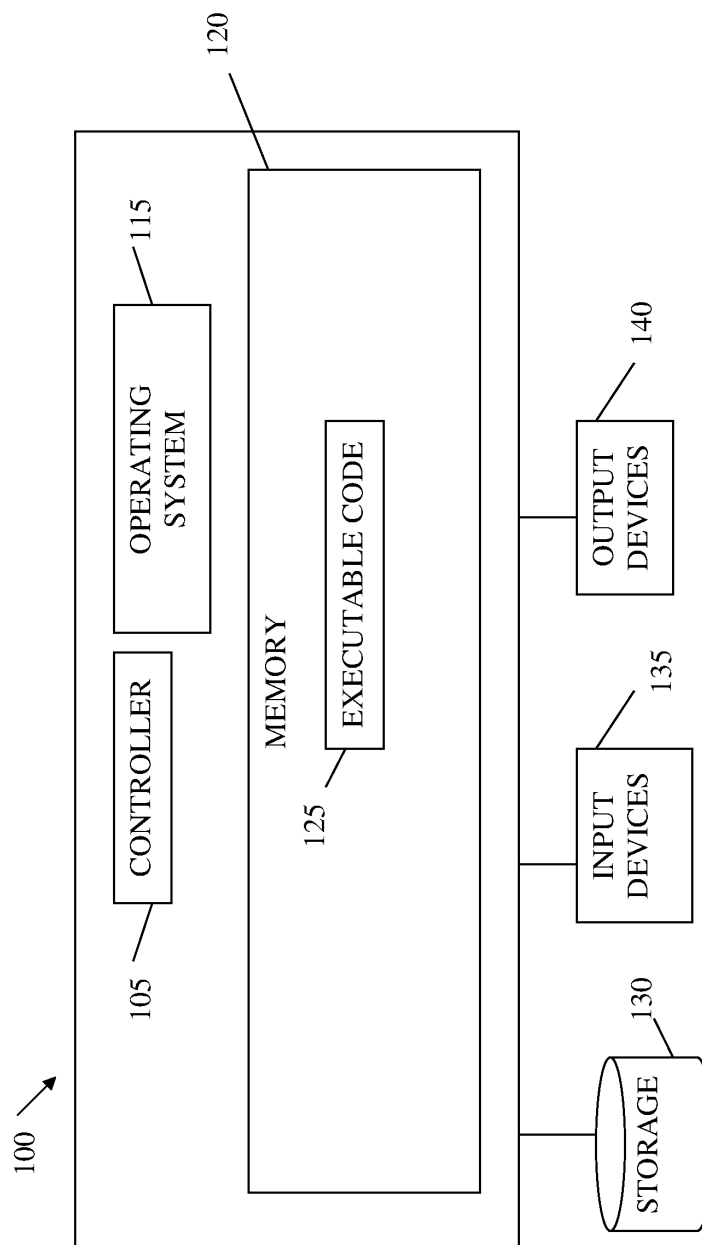


Fig. 2

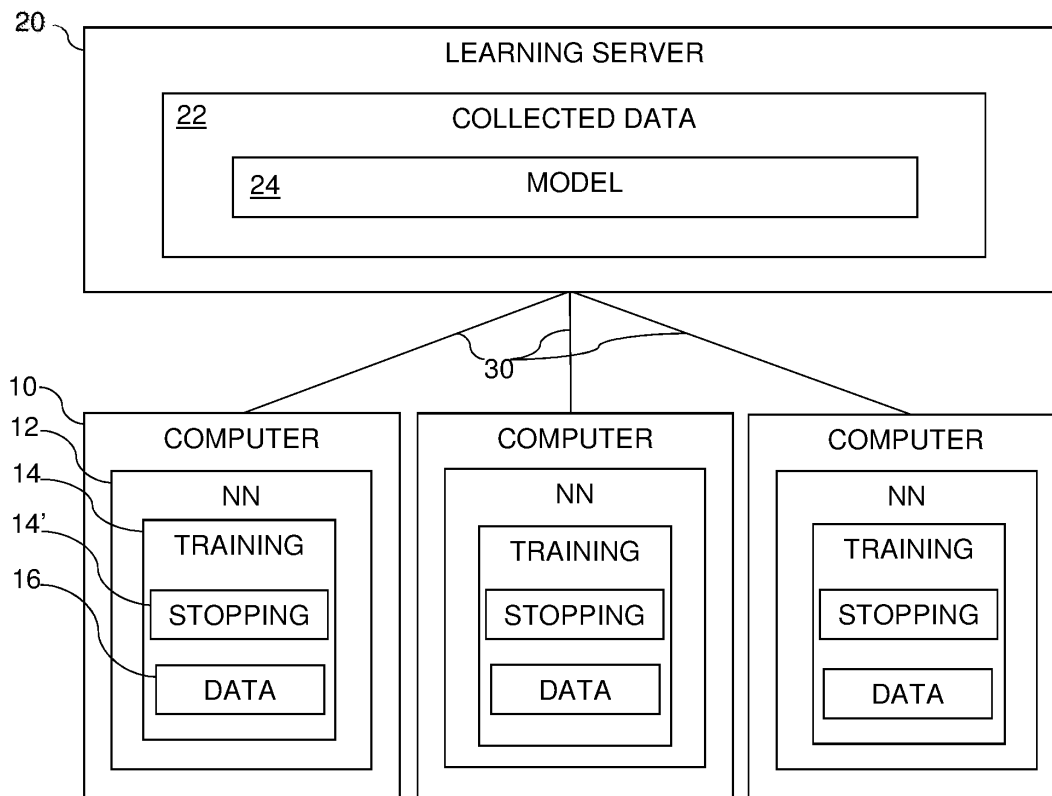


Fig. 3

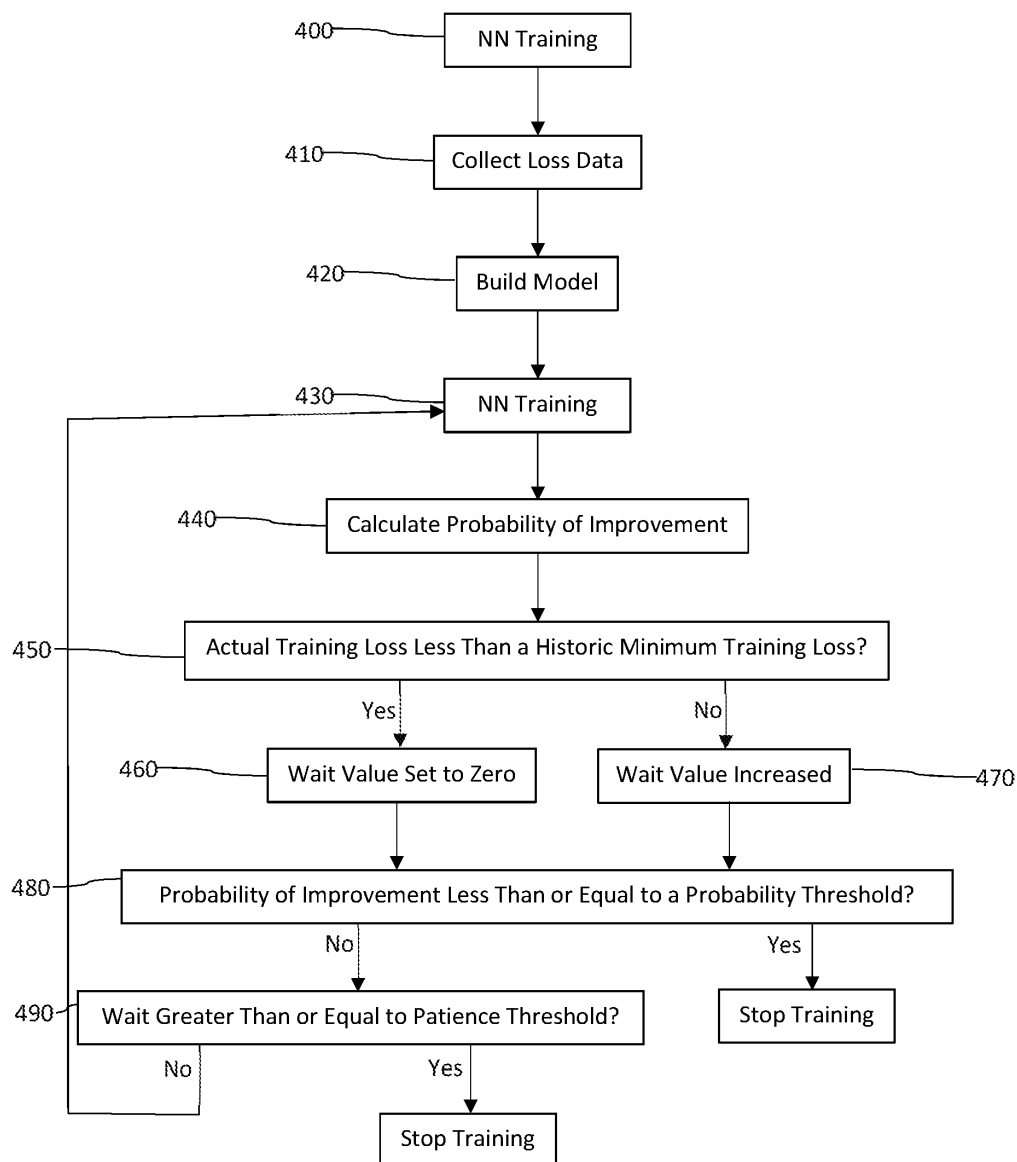


Fig. 4A

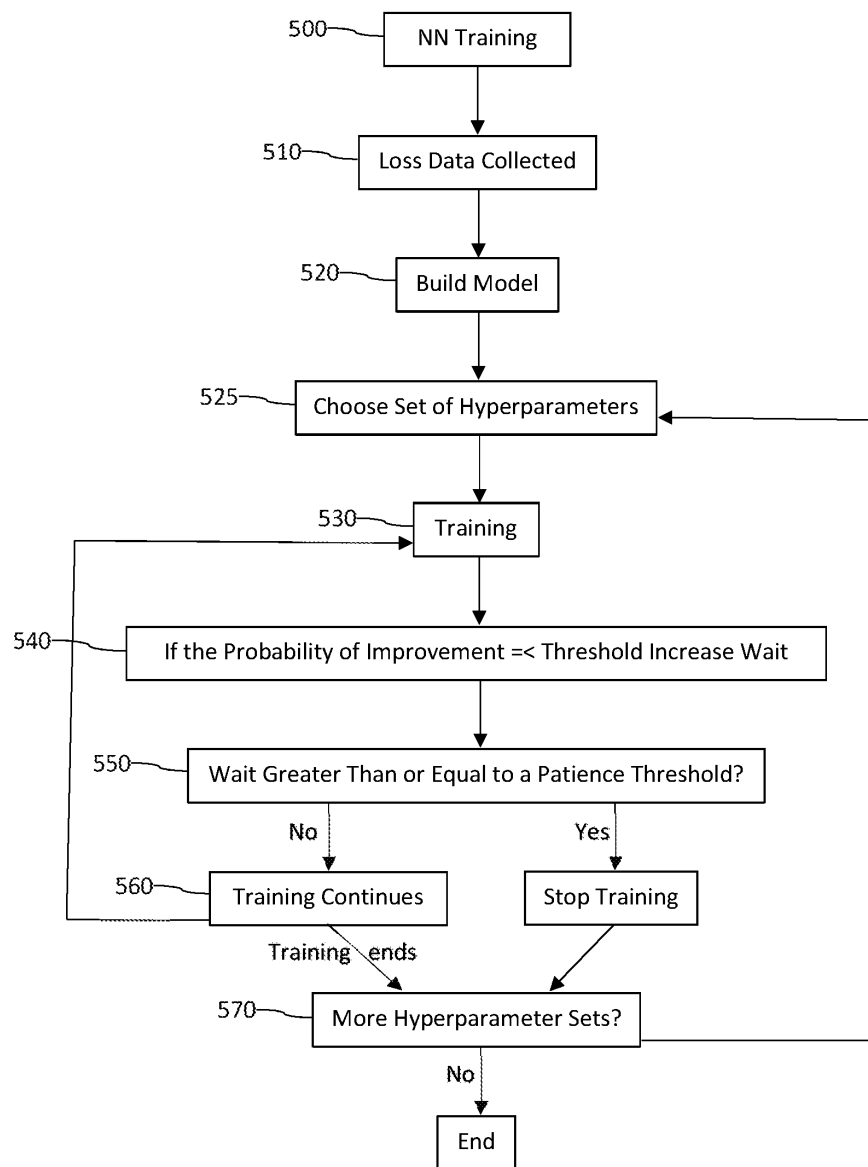


Fig. 4B

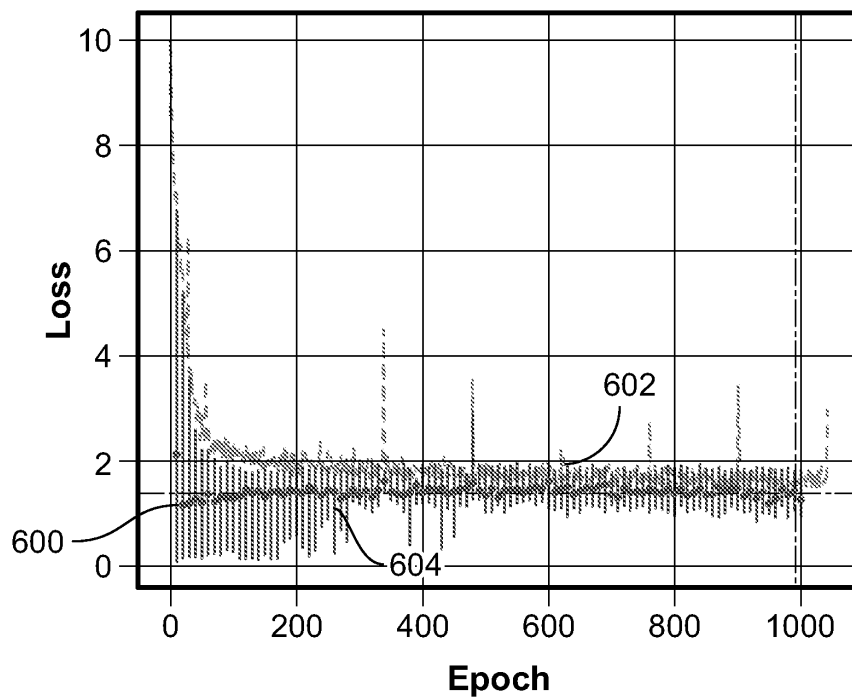


Fig. 5A

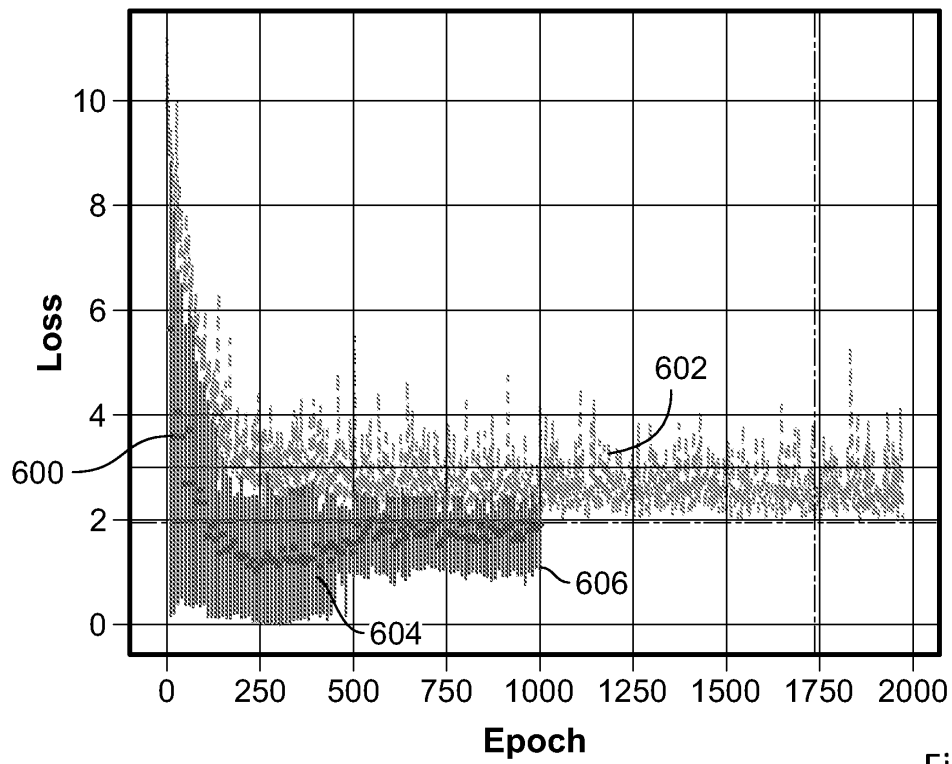


Fig. 5B

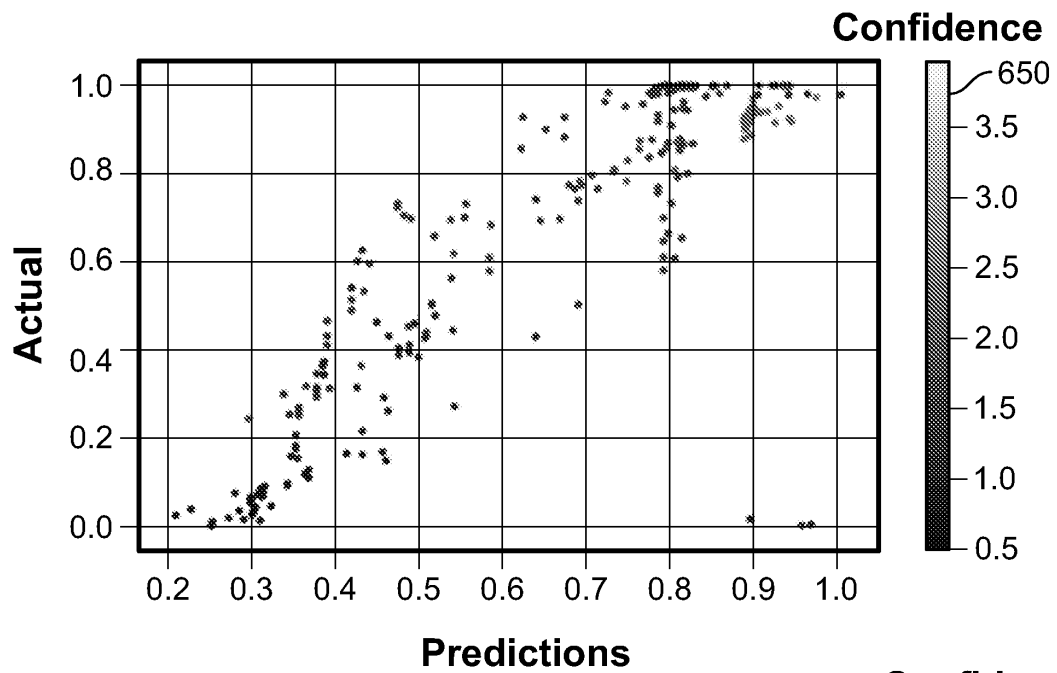


Fig. 6A

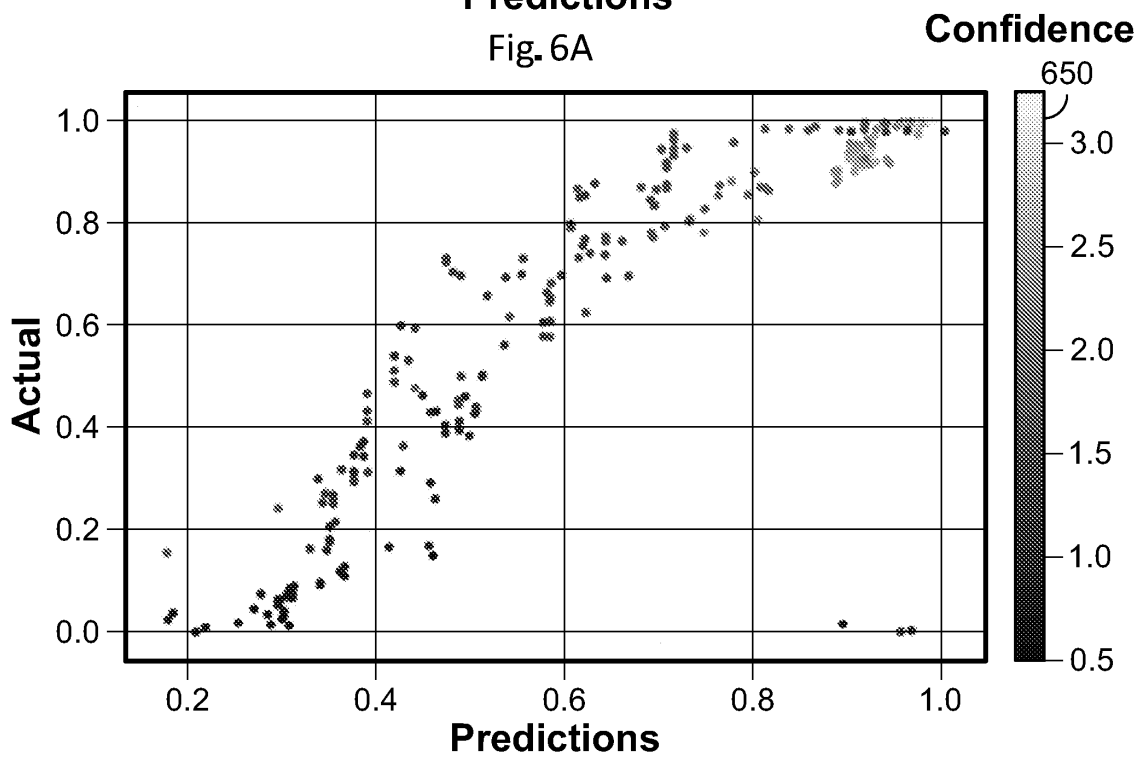


Fig. 6B

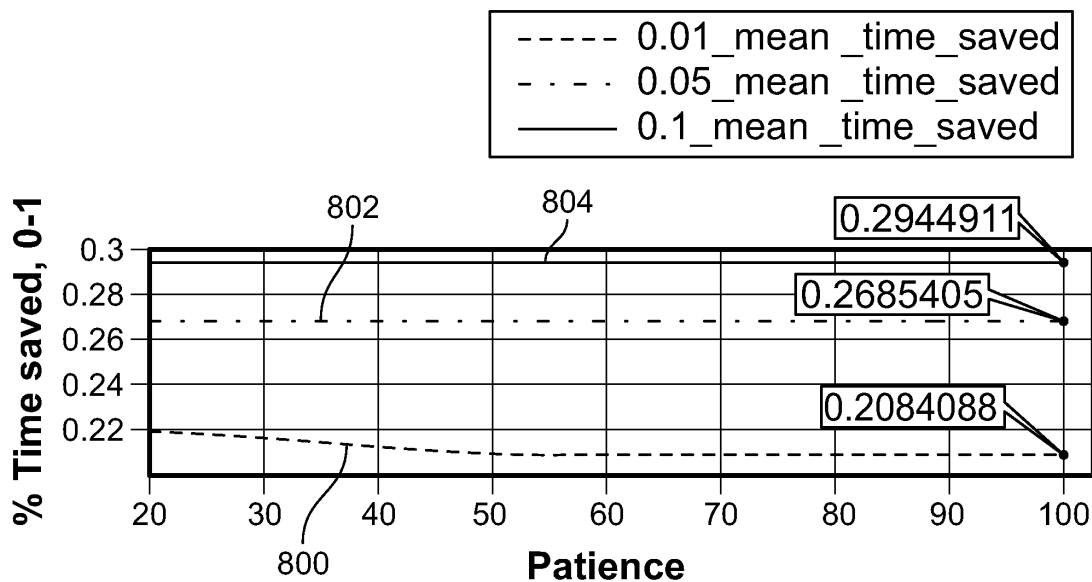


Fig. 7A

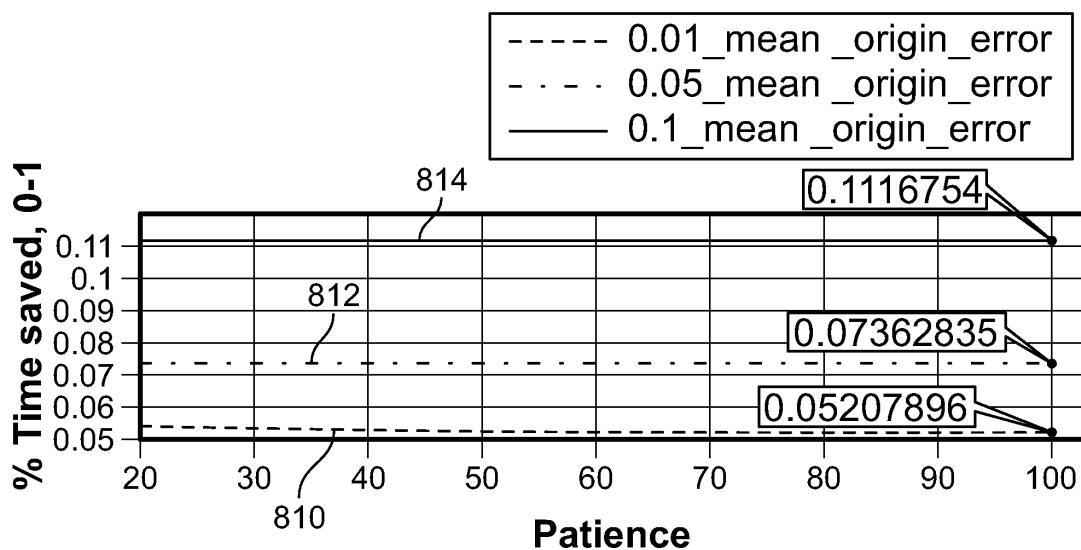


Fig. 7B

ABSTRACT

Systems and methods may train neural networks (NNs) and determine when to stop training to not waste computing or other resources when improvement is not no longer likely. After training period for a NN, a model trained using training data from other NNs may return a probability of improvement in the loss of the NN or a probability that the likely best loss of the NN is lower than the best loss of the other NNs for which hyperparameters have been chosen. Training may be stopped if the probability is less than a threshold, or a wait value is greater than a wait threshold.

What is claimed is:

1. A method of training a neural network (NN) having a first set of hyperparameters after having trained a set of other NNs each having a set of hyperparameters different from the first set of hyperparameters, the method comprising:
over a series of NN training epochs, where in each epoch the NN undergoes training and a loss is computed:
determining, using:
a set of model parameters;
data describing training loss of the NN; and
a model that has been trained using training losses of a plurality of NNs other than the NN;
a probability that the likely best loss of the NN is lower than the best loss of the other NNs; and
if the probability is less than a threshold, or a wait value is greater than a wait threshold, stopping training.
2. The method of claim 1, comprising if the probability is not less than a threshold and a wait value is not greater than a wait threshold, continuing training.
3. The method of claim 1, comprising increasing the wait value if the probability is less than or equal to a probability threshold.
4. The method of claim 1, wherein determining using a model a probability that the likely best loss of the NN is lower than the best loss of the other NNs comprises obtaining from a leaf of at least one tree data structure data relevant to the NN.
5. The method of claim 1, wherein determining a probability comprises determining a mean expected training loss and a variance.
6. The method of claim 1, wherein model parameters comprise hyperparameters.

7. A method of training a neural network (NN), the method comprising:
over a series of NN training epochs, where in each epoch the NN undergoes training and a loss is computed:
determining, using:
a set of model parameters;
data describing training loss of the NN; and
a model that has been trained using training losses of a plurality of NNs other than the NN;
a probability of improvement in the loss of the NN; and
if the probability is less than a threshold, or a wait value is greater than a wait threshold, stopping training.
8. The method of claim 7, comprising if the probability is not less than a threshold and a wait value is greater than a wait threshold, continuing training.
9. The method of claim 7, comprising increasing the wait value if the current loss of the NN is not less than the minimum loss in the loss history for the NN.
10. The method of claim 7, comprising setting the wait value to zero if the current loss of the NN is less than or equal to the minimum loss in the loss history for the NN.
11. The method of claim 7, wherein determining using a model a probability of improvement in the loss of the NN comprises obtaining from a leaf of at least one tree data structure data relevant to the NN.
12. The method of claim 7, wherein determining an expected probability of improvement comprises determining a mean expected training loss and a variance.
13. The method of claim 7, wherein model parameters comprise hyperparameters.

14. A system for training a neural network (NN) having a first set of hyperparameters after having trained a set of other NNs each having a set of hyperparameters different from the first set of hyperparameters, the system comprising:
- a memory; and
 - a processor configured to:
 - over a series of NN training epochs, where in each epoch the NN undergoes training and a loss is computed:
 - determine, using:
 - a set of model parameters;
 - data describing training loss of the NN; and
 - a model that has been trained using training losses of a plurality of NNs other than the NN;
 - a probability that the likely best loss of the NN is lower than the best loss of the other NNs; and
 - if the probability is less than a threshold, or a wait value is greater than a wait threshold, determine to stop training.
15. The system of claim 14, wherein the processor is configured to, if the probability is not less than a threshold and a wait value is not greater than a wait threshold, determine to continue training.
16. The system of claim 14, wherein determining using a model a probability that the likely best loss of the NN is lower than the best loss of the other NNs comprises obtaining from a leaf of at least one tree data structure data relevant to the NN.
17. The system of claim 14, wherein determining a probability comprises determining a mean expected training loss and a variance.
18. A system of training a neural network (NN), the system comprising:
- a memory; and
 - a processor configured to:

over a series of NN training epochs, where in each epoch the NN undergoes training and a loss is computed:

determine, using:

a set of model parameters;

data describing training loss of the NN; and

a model that has been trained using training losses of a plurality of NNs other than the NN;

a probability of improvement in the loss of the NN; and

if the probability is less than a threshold, or a wait value is greater than a wait threshold, determine to stop training.

19. The system of claim 18, wherein the processor is configured to, if the probability is not less than a threshold and a wait value is greater than a wait threshold, determine to continue training.
20. The system of claim 18, wherein the processor is configured to increase the wait value if the current loss of the NN is not less than the minimum loss in the loss history for the NN.
21. The system of claim 18, wherein the processor is configured to set the wait value to zero if the current loss of the NN is less than or equal to the minimum loss in the loss history for the NN.
22. The system of claim 18, wherein determining using a model a probability of improvement in the loss of the NN comprises obtaining from a leaf of at least one tree data structure data relevant to the NN.

SYSTEMS AND METHODS FOR PREDICTIVE EARLY STOPPING IN NEURAL NETWORK TRAINING

RELATED APPLICATION DATA

[0001] This application claims benefit from US provisional patent application 62/852,525, filed on May 24, 2019 and entitled “SYSTEM AND METHOD FOR PREDICTIVE EARLY STOPPING IN NEURAL NETWORKS” incorporated herein by reference in its entirety.

FIELD OF THE INVENTION

[0002] The present invention relates generally to machine learning; specifically to training neural networks.

BACKGROUND

[0003] NNs are computing systems inspired by biological computing systems, but operating using manufactured digital computing technology. NNs are made up of computing units typically called neurons (which are typically artificial or simulated neurons, as opposed to real or biological neurons) communicating with each other via connections, links or edges. In common NN implementations, the signal at the link between artificial neurons may be for example a real number, and the output of each neuron may be computed by function of the (typically weighted) sum of its inputs. NN links or edges typically have a weight that is adjusted as learning proceeds. The weight increases or decreases the strength of the signal at a connection. Typically, NN neurons are divided or arranged into layers, where different layers may perform different kinds of transformations on their inputs and may have different patterns of connections with other layers. NN layers may also be connected by functions which may be adjusted with learning.

[0004] Machine learning algorithms, using e.g. neural network (NN) or connectionist systems, are trained or learn by iteratively adjusting model parameters to optimize an objective function, also known as a loss function. The goal of training a model may be to determine a set of hyperparameters, and/or weights of links or functions connecting neurons, and an optimization procedure that minimizes the training loss and allows the model to achieve generalization on unseen data.

[0005] Hyperparameters may be parameters that cannot be directly optimized as a function of the model's loss, and may be determined using a part of the data that has not been used for training (e.g. a validation or hold out set). NN hyperparameters may describe an overall architecture of the NN such as number and structure of layers, number of neurons in each layer or overall, the type

of functions neurons use when computing output, and variables which determine how the network is trained, such as the learning rate. Hyperparameters may include parameters other than those describing a NN architectures, for example NN training parameters such as batch size, number of training parameters, number of training samples, number of link weights, number of layers, number of examples in a training set, etc. Hyperparameters may include for example, a number of layers, a kernel size for a layer (e.g. 3x3, 7x7), a batch size (e.g. 32, 64, 128 or 256), a learning rate (e.g. 0.0001 to 0.1) or other parameters. Hyperparameters may include layer configuration, activation function, or other data. Other hyperparameters may be used.

[0006] A NN may also be described by non-hyperparameter parameters such as weights of links or functions that link layers of a NN, which may be adjusted during learning or training.

[0007] Training a model, such as large NN models, is a time-consuming process, with a significant computational overhead. Evaluating different weights, and hyperparameter configurations, involves waiting for the model's loss function to converge to some local minima.

[0008] NNs may learn to perform tasks by considering example input data, generally without being programmed with any task-specific rules, being presented with the correct output for the data, and self-correcting. During learning the NN may execute a forward-backward pass where in the forward pass the NN is presented with an input and produces an output, and in the backward pass (backpropagation) the NN is presented with the correct output, generates an error e.g., a loss, and generates update gradients which are used to alter the weights at the links or edges. Different types of NNs may generate different types of outputs (e.g. binary, a numerical range) and different numbers of outputs.

[0009] During learning, the NN, or the computing nodes simulating the NN, may be presented with training data. For example, in an image recognition application, a NN may learn to identify images that contain cats by analyzing example images that have been manually labeled as "cat" or "not a cat" and using the results to identify cats in other images. The NN may do this without any prior knowledge about cats, e.g., that they have fur, tails, whiskers and cat-like faces. Instead, during learning the NN automatically generates identifying characteristics from the learning material that it processes.

[0010] In practice, a NN, and NN learning, is simulated by one or more computing nodes, such as CPUs, e.g. as embodied in personal computers or graphics processing units (GPUs) such as provided by Nvidia Corporation. One or more computing nodes may simulate a NN using known

data structures. During execution or inference, the trained NN may for example recognize or categorize images, perform speech processing, or other tasks. A NN may be simulated as an abstract mathematical object, such as a function. A NN may be translated physically to CPU or GPU as for example a sequence of matrix operations where entries in the matrix represent neurons (e.g. artificial neurons connected by edges or links) and matrix functions represent functions of the NN.

[0011] A loss or function may define a deviation from an expected or correct output value for a given input, during learning or training. A loss function for a network may represent the difference or inconsistency between the value or values output from the network, and the correct value/values that should be output given the data input to the NN. A loss function may be, for example, a negative log-likelihood or residual sum of squares, but may be computed in another manner. In NN learning, it is desired to minimize loss, and after receiving a loss the NN representation may be updated by modifying weight values in the network using backpropagation.

[0012] Learning or training may be organized in batches and epochs. A batch may be a set of multiple training data presented to a NN which results in a number of outputs. The NN weights may be modified after an entire batch is run. A loss may be generated for each batch. An epoch may be a set of batches representing the entire training set; thus an epoch may represent a number of forward/backward passes, each resulting in weight adjustment. There may be multiple losses per epoch, one for each batch. For example, if there are 1,000 samples of training data, divided into 100 batches of 10 data samples each, the training using all 100 batches may make up one epoch. A NN may be trained in multiple epochs, each epoch using the same training data as the last.

[0013] Prior training methods typically stop by periodically testing a NN using a holdout data set not included in training data. Training may be stopped for example when improvement stagnates. Since training is a lengthy process, it is desirable to reduce the time spent training. Training NN often involves a diminishing amount of improvement over time, and it is desirable to determine at what point training can be stopped if an acceptable amount of improvement is achieved, without wasting time on further training due to the law of diminishing returns. It is hard or impossible using prior art methods to predict at any given point in training how much improvement can be achieved by training using further epochs. It would be desirable to determine at what point further

improvement beyond a certain amount is not likely, or not worth the extra time and resources for training.

SUMMARY

[0014] Systems and methods of the present invention may train one or more neural networks (NNs) and determine when to stop training to not waste computing or other resources when improvement is not no longer likely. Systems and methods may train neural networks (NNs) and determine when to stop training to not waste computing or other resources when improvement is not no longer likely. After training period (e.g. an epoch) for a NN, a model trained using training data from other NNs may return a a probability of improvement in the loss of the NN or a probability that the likely best loss of the NN is lower than the best loss of the other NNs for which hyperparameters have been chosen. Training may be stopped if the probability is less than a threshold, or a wait value is greater than a wait threshold.

[0015] In one embodiment, in a “global” mode, training or evaluating a neural network (NN) which has a first set of hyperparameters, and where there are other trained NNs each having a set of hyperparameters different from the first set of hyperparameters, may include over a series of NN training periods or epochs, where in each epoch the NN undergoes training and a loss is computed, determining a probability that the likely best loss of the NN is lower than the best loss of the other NNs. The determining may be performed using one or more models (which may include sub-models) that has been trained using training losses of a plurality of NNs other than the NN, with different sets of hyperparameters and numbers of epochs used for training as the NNs the current NN is being compared to for training) having input to it a set of model parameters and data describing training loss of the NN. If the probability is less than a threshold, or a wait value is greater than a wait threshold, training may be stopped. In global mode, the model may predict the probability that the current loss curve will be able to improve beyond the best performing model (e.g. NN with certain hyperparameters) seen so far. In local mode, the model may predict the probability that the NN will improve beyond the current minimum value of the loss curve.

[0016] In one embodiment, in a “local” mode, training or evaluating a neural network (NN) may include over a series of NN training epochs, where in each epoch the NN undergoes training and a loss is computed, determining, using a model (or one or more models; models may include sub-models), a probability of improvement in the loss of the NN. The model may have input to it a set of model parameters and data describing training loss of the NN. The model(s) may have been

trained using training losses of a plurality of NNs other than the NN. If the probability is less than a threshold, or a wait value is greater than a wait threshold, training may be stopped.

[0017] In one embodiment, over a series of NN training epochs, where in each epoch the NN undergoes training and a loss is computed, an expected training loss may be determined, using a set of model parameters, data describing training loss of the NN, and a model that has been trained using training losses of a plurality of NNs other than the NN. If the expected training loss (possibly less or minus a loss threshold) is greater than or equal to an actual training loss, or if a wait value is greater than a wait threshold, training may be stopped; otherwise training may continue.

[0018] One embodiment of a predictive early stopping tool may estimate the value of a local minima beforehand, and may terminate the model training once it is achieved. Example results show embodiments may reduce model training time by an average of 20%, with an average error rate of 4%.

BRIEF DESCRIPTION OF THE DRAWINGS

[0019] The patent or application file contains at least one drawing executed in color. Copies of this patent or patent application publication with color drawing(s) will be provided by the Office upon request and payment of the necessary fee.

[0020] Non-limiting examples of embodiments of the disclosure are described below with reference to figures attached hereto that are listed following this paragraph. Dimensions of features shown in the figures are chosen for convenience and clarity of presentation and are not necessarily shown to scale.

[0021] The subject matter regarded as the invention is particularly pointed out and distinctly claimed in the concluding portion of the specification. The invention, however, both as to organization and method of operation, together with objects, features and advantages thereof, can be understood by reference to the following detailed description when read with the accompanied drawings. Embodiments of the invention are illustrated by way of example and not limitation in the figures of the accompanying drawings, in which like reference numerals indicate corresponding, analogous or similar elements, and in which:

[0022] Fig. 1 is a block diagram of a neural network according to an embodiment of the present invention.

[0023] Fig. 2 is a high-level block diagram of an exemplary computing device which may be used with embodiments of the present invention.

[0024] Fig. 3 is a high-level block diagram of an exemplary system which may be used with embodiments of the present invention.

[0025] Fig. 4A is a flowchart of a method according to embodiments of the present invention.

[0026] Fig. 4B is a flowchart of a method according to embodiments of the present invention.

[0027] Figs. 5A and 5B depict example prediction loss curves, according to one embodiment of the invention.

[0028] Figs. 6A and 6B depict actual changes in loss values compared with predicted changes in the loss curves, according to one embodiment of the invention.

[0029] Figs. 7A and 7B depict example early stopping results, according to one embodiment of the invention.

[0030] It will be appreciated that for simplicity and clarity of illustration, elements shown in the figures have not necessarily been drawn accurately or to scale. For example, the dimensions of some of the elements can be exaggerated relative to other elements for clarity, or several physical components can be included in one functional block or element.

DETAILED DESCRIPTION

[0031] In the following detailed description, numerous specific details are set forth in order to provide a thorough understanding of the invention. However, it will be understood by those skilled in the art that the present invention can be practiced without these specific details. In other instances, well-known methods, procedures, and components, modules, units and/or circuits have not been described in detail so as not to obscure the invention.

[0032] Embodiments of the invention include systems and methods that may stop training of a NN if it is determined that likelihood of improvement in loss over a best loss seen is below a threshold, or the training has proceeded to the mean expected or final value of training (typically after a certain number of initial training cycles or epochs), or if training has “stalled” and does not improve (e.g. over a historic best value) over a certain number of training cycles or epochs. In some embodiments, a training module or unit may call an early stopping module or unit (e.g. as described in the examples in Tables 1-5 herein) and receive instructions to stop training or not stop training.

[0033] Embodiments may be used for choosing hyperparameters of a NN. For example, hyperparameters may be chosen for a NN as is known in the art, and the NN may be trained, the training process determining if to stop training the NN if the likelihood of improvement (e.g. to a final loss) of the NN beyond a best loss seen for other NNs with other sets of hyperparameters is less than a threshold. In another embodiment, after hyperparameters for a NN are decided, the NN may be trained by a process which stops training when the likelihood of improvement (e.g. to a final loss) of the NN beyond a best loss seen before for this NN is less than a threshold. Other embodiments may stop training based on a comparison of the current loss for the NN to a predicted best loss (e.g. final best loss) for the NN. In one embodiment, if an expected or predicted training loss (e.g. that returned by a model) is greater than or equal to the actual training loss of the target NN, or if patience has been exhausted (e.g., if a wait value, which is incremented if no improvement is seen, is greater than a wait threshold) training may be stopped.

[0034] In some embodiments, one or more models, possibly including sub-models, may be built or trained using training data of many other NNs (e.g. collected in the past, before training the target NN, or before choosing hyperparameters for a set of NNs to be compared), some having different structures and uses than the target NN being trained, and the model may be used to return an expected loss at a certain time or training period for the target NN. The model may take as input data describing the NN being trained such as model parameters (e.g. hyperparameters) and loss data such as a loss history and a current loss (which may be included in the loss history), and return one or more values describing a training loss.

[0035] Typically, there is a delay before determining whether or not to stop to prevent stopping from occurring at the very beginning of training in an early training cycle, period or epoch (e.g. the first or second epoch). For example, a process may wait for a certain number of iterations or epochs, e.g. 10 epochs, for a target NN being trained, before making a prediction and determining if training should be stopped. Additionally, the prediction may be adjusted for every new epoch of data received, and thus the mean predicted loss value may rise or fall based on loss values being reported.

[0036] Embodiments may incorporate a predictive model into an early stopping algorithm that is responsible for stopping a NN training process at the point at or as soon as the minimum predicted validation loss is achieved. An embodiment may terminate training based on a model, such as a prediction provided by a learning curve model.

[0037] The model may be updated using data gained from training the particular or target NN. The comparison of expected and actual loss may be modified by a threshold or range such that if the difference is less than the threshold or within the range it can be deemed that stopping should occur.

[0038] Some embodiments may be built on top of a model estimating the minimum value of a validation metric for a given experiment configuration. A large amount of loss function data, or training losses, may be collected, from multiple different NNs (typically different NNs in a variety of different contexts, with a variety of different architectures, performing a variety of different analysis tasks, having been trained for different numbers of epochs), along with hyperparameter and possibly other data, e.g. via an experiment or learning management platform, to build or train a predictive model that can estimate an interval in which an evaluation metric is likely to fall. For example, time series data or training losses from thousands of loss curves, from many different NNs, may be extracted from experiments or training executed using an experiment management platform. In addition to the time series data, derived data or parameters may be extracted for each curve of loss over time, such as slope and rate of change. Some embodiments may, unlike some prior art methods, not partition data based on specific NN architectures. Some embodiments do not need to use information regarding architecture type, or architecture specific parameters, such as the number of convolution filters, to predict the minimum value of the learning curve.

[0039] In some embodiments, since the models used have been trained or developed using training losses of multiple NNs, a particular NN when applied to the model to determine stopping gets the benefit of a model trained on data gathered from many NNs including those other than that particular NN being analyzed for stopping.

[0040] The model or models may be gradient boosting machines (GBNs). Gradient boosting is a machine learning technique for classification, where a prediction model may be in the form of an ensemble of weak prediction models, typically decision trees including, e.g. leaves which group NNs into groups having parameters falling within ranges of hyperparameters or other descriptive data. Other ways of structuring a model may be used. If the models are trees, in each tree, the NNs may be grouped or organized differently, and thus in each tree a given NN may fall into a different leaf. Typically a tree used for a model is initialized with a fixed leaf structure, each leaf accepting and storing data for a set of NNs having a certain range of parameters. Data describing an NN may be input to a model including a tree structure, a leaf representing or including that NN may

be found, and data relevant to the NN (e.g. a mean expected loss, a variance of loss, a range of expected loss, a probability of improvement) may be obtained from the leaf. Typically, the output values for a model are created from the average of the relevant values for the relevant leaf. In one embodiment, a model may include an individual tree including data for each specific epoch or time period; in another embodiment an individual tree may include data for NNs having been trained for different numbers of epochs or time periods, and specific leaves may describe NNs having been trained using different numbers of epochs.

[0041] In one embodiment, early stopping may be based on a calculated probability that loss will improve, the probability generated by a model categorizing NNs into leaves or other buckets or sets, based on NN characteristics. In such an embodiment, NNs may be characterized on derivations or features of loss histories, and in some embodiments only two “tree” models need be used, with no separate models based on the epoch. For example, in some embodiments, one model or tree may categorize NNs having been trained with varying numbers of epochs, and a single set or leaf may contain NNs having been trained over different numbers of epochs. In such a model, a single NN may appear in multiple places in the same tree, as its features or characteristics may change over time, for example with additions to its loss history.

[0042] In one embodiment, a first model (e.g. m_μ) may return, based on an input of NN characteristics or features (e.g. ux), a mean best (e.g. final) expected loss for that NN; and a second model (e.g. m_σ^2) may return, based on that NN’s characteristics or features, the variance around the mean (e.g. an estimate for the uncertainty of the associated mean, e.g. the mean for the NN with those specific characteristics) for the NN having those specific characteristics. Each of the models may return data relevant to the NN, which may be used to compute a probability of improvement. In an early epoch a variance may be high, and that variance may become smaller the more epochs exist. Typically m_μ and m_σ^2 take as inputs the same characteristics and the variance returned by m_σ^2 may be associated with or correspond to the mean of m_μ . The variance may correspond to the span indicating the best case and worst case for the loss, with the mean at the center, returned by m_μ . One embodiment may estimate the minimum expected convergence value of a loss function, $y(t)$, given a feature set of model hyperparameters, ux , and time series data of the loss curve.

[0043] In one embodiment, a model, or one or more separate models, e.g. tree-based regression models, may be trained for every training cycle, period or epoch of time series data in a dataset,

such that each leaf in each model typically includes data regarding NNs having been trained on different number of epochs. The models (e.g. m_μ and m_δ^2) may each be trained using separate or non-overlapping data. For example, the various data points of NNs at various stages of learning (e.g. having different numbers of datapoints in a loss history) may be divided up, and half may be used to train m_μ and half may be used to train m_δ^2 . In one embodiment, a model producing mean expected loss (e.g. m_μ) may first be trained using dataset X (half of the data) and thus may provide an expected convergence value. Then that first model may be used to make predictions about the minimum convergence value on the second half of the dataset, and a model producing variance (e.g. m_δ^2) may be trained and having data input by inputting data Y (the other half of the data) to that first model and comparing actual loss for the specific NNs with the mean from the first model: the resulting variance may be used to train and input data into a second model. The second model (e.g. m_δ^2) may be trained to predict the squared residual (e.g. error or variance) values of the predictions of the first model (e.g. m_μ) using the same features as m_μ . Thus m_μ may provide an estimate of the mean value convergence, while m_δ^2 may provide an estimate of the variance. This may allow the creation of confidence around predicted values. Thus a combined model used may be:

$$f(x) = \{m_\mu, m_\delta^2\}$$

[0044] In another embodiment, three separate models, e.g. tree-based regression models, may be trained for every training cycle, period or epoch of time series data in a dataset. The models may each return (e.g. by finding leaves corresponding to the NN) data relevant to the NN. The first model may provide estimates of the mean or average expected change in the loss curve. The second and third models may use a quantile loss that estimates a lower percentile and an upper percentile, e.g. the 5th and 95th quantile change, in the loss curve. For example, the low model may return a value indicating the loss such that X% (e.g. 5%) of the NNs matching the input (e.g. loss history, hyperparameters) have a loss higher or equal to the value. The high model may return a value indicating the loss such that X% (e.g. 95%) of the NNs matching the input (e.g. loss history, hyperparameters) have a loss higher or equal than the value. This may create, for example, a 90% prediction interval for an estimated loss value.

[0045] In one embodiment, a final model may include a set of individual models which may be responsible for predicting the minimum value of convergence based on the current training epoch, such as the following example:

$$f(x) = \{\{m_1^{10}(x^{10}), m_2^{10}(x^{10}), m_3^{10}(x^{10})\}, \dots, \{m_1^t(x^t), m_2^t(x^t), m_3^t(x^t)\}\}$$

Where m_1^X is a first model, returning an estimated mean, for epoch or other period of time X (and which may be considered to have X sub-models each corresponding to an epoch or other period of time), m_2^X is a second model, returning an upper or high bound for an estimated range (a quantile loss that estimates an upper percentile), for epoch X; and m_3^X is a third model, returning a lower or low bound for an estimated loss (a quantile loss that estimates a lower percentile), for epoch X. In one embodiment, a wait period is applied such that the model is not applied until the X'th (e.g. 10th) epoch is used, as typically loss will decrease during initial training epochs and premature stopping due to actual loss being greater than estimated loss is not desired during this early period. In such a model each leaf describes NNs having been trained on the same number of epochs.

[0046] In one embodiment, a model may include multiple “sub-models”: a single model is responsible for returning mean loss data at a certain epoch (where the epoch may be measured by the number of values in a series of loss data); another model is responsible for returning an upper or high limit for a range at that epoch, and a third model is responsible for returning a lower limit for the range at that epoch. In some models, a set of models may each include sub-models, each sub-model corresponding to NNs having been trained over a certain number of epochs or to NNs having a certain numbers of losses in their loss history. In other embodiments, a model may include in the same categories or leaves NNs having different numbers of losses in their loss histories. For a given hyperparameter configuration, the models (e.g. models m_1 , m_2 and m_3 in the example above) may be responsible for predicting the minimum value of the loss curve based on, for example, 100 epochs worth of data, or other volumes of data.

[0047] Time series features used as input to a model may be loss values or loss functions, or validation loss functions from training or experiments. Typically, loss data or time series data is the loss recorded for the last loss in the last batch in each epoch over a series of epochs, and thus such data is organized by epoch. Typically, one loss is generated per batch, and multiple losses are generated per epoch, one per each batch. E.g., during an epoch, a number of losses may occur, but the one used for input to a model may be the last loss occurring in the epoch. Other of different ways of calculating a series of loss data, and defining which loss values to use, may be used. In some embodiments, learning may not be organized by batch or epoch.

[0048] In some embodiments, for any model in the set of models used, responsible for making a prediction based on the n th epoch of data, the raw time series features may be processed. For example, since different NNs have different formats and ranges of loss values, loss values (e.g. for any n 'th epoch slice of data) may be normalized or rescaled, e.g. by applying a log transformation, or to the scale 0-1, or other scales. Typically, multiple loss values for a NN are normalized to one value: if a given NN returns a number of loss values per epoch, the multiple loss values may be converted to a single loss value for use with a model. For example, raw time series features or loss values may be first smoothed, for example by applying a log transformation, followed by a Minmax scaling.

[0049] After preprocessing (e.g. normalizing, smoothing, etc.) certain features of the values may be extracted or calculated and used in models to categorize instances of NN behavior, e.g. to categorize NN instances as leaves in a tree model. The features extracted may be used to decide into which leaf or category to place the NN. Extracted time series features may be input along with hyperparameters and graph features to an alignment process. Hyperparameters and other NN characteristics may be extracted from for example a graph representing the NN (a NN may be represented as a graph of nodes connected by links). Extracted timeseries features, hyperparameters (e.g. batch size, learning rate, number of training samples, etc.), and graph (e.g. extracted from the graph representing the NN) or other features such as the number of trainable parameters present in the mode may be analyzed. Example values or features that can be extracted to describe training loss include:

- First and second order difference of the normalized or pre-processed time series;
- Summary statistics of the first and second order differential features, such as the mean, median, sum, standard deviation, etc.
- Autocorrelation coefficients of the raw, untransformed time series data. The mean and standard deviation of this set of data may be used as features. Autocorrelation coefficients of a time series may provide a measure of how strongly a value in the time series correlates to a lagged version of itself. For example, when a lag value of 1 is used, how strongly correlated the time series at $t-1$ is to the time series at t may be calculated. A lag of 1 will result in a single autocorrelation coefficient. In some embodiments, a lag of, for example, up to 40 may be used, resulting in a maximum of

40 coefficients. The mean and standard deviation of these coefficients may be used as features.

The raw time series features used to create the derived data may be dropped or ignored, or may be set to “not a number”, or NAN. Other or different features may be used.

[0050] In some NNs, during backpropagation, each neuron computes its own gradient for a link for the neuron, the gradient to be applied to adjust the weight of the link. When discussed herein, a neuron taking action such as transmitting data, computing data, etc., may mean that a processor simulating the neuron performs a computation to simulate such action; e.g. a computing node simulating a number of neurons may perform the actual action that is ascribed to the neuron. Typically parameters such as weights are represented as floating point (e.g. 32 bit) numbers, but may be represented in other ways, such as integers or numbers represented by different numbers of bits.

[0051] Fig. 1 is a simplified block diagram of a NN used with an embodiment of the present invention; in typical use thousands of neurons and links are used. NN 1000 may input data as for example an input vector 1010 of values (representing, e.g. a photograph, voice recording, or any sort of data), and may produce an output of signals or values, for example a single value or output vector 1020. NN 1000 may have neurons arranged into layers 1030, each including neurons 1040 connected to other neurons by links or edges 1050. NNs may be used for different tasks. NNs used for classification tasks, e.g. classifying photographs into descriptions of the content, may produce, for each class i , an output z_i , sometimes called a logit, which may encode or represent the likelihood that a given example input should be classified to class i . Logits z_i , for each class i , (e.g., for image recognition dog, cat, llama, etc.) may be transformed into probabilities q_i by comparing each z_i to the other logits, in for example a softmax layer.

[0052] Fig. 2 shows a high-level block diagram of an exemplary computing device which may be used with embodiments of the present invention. Computing device 100 may include a controller or processor 105 that may be or include, for example, one or more central processing unit processor(s) (CPU), one or more Graphics Processing Unit(s) (GPU or GPGPU), a chip or any suitable computing or computational device, an operating system 115, a memory 120, a storage 130, input devices 135 and output devices 140. Each of modules and equipment discussed elsewhere (e.g. Fig. 3) such as server 20, model(s) 24, computers 10, NNs 12, training software 14, stopping module 14', and other equipment and modules mentioned herein may be, include, or

be executed by a computing device such as included in Fig. 2, although various units among these entities may be combined into one computing device.

[0053] Operating system 115 may be or may include any code segment designed and/or configured to perform tasks involving coordination, scheduling, arbitration, supervising, controlling or otherwise managing operation of computing device 100, for example, scheduling execution of programs. Memory 120 may be or may include, for example, a Random Access Memory (RAM), a read only memory (ROM), a Dynamic RAM (DRAM), a Synchronous DRAM (SD-RAM), a double data rate (DDR) memory chip, a Flash memory, a volatile memory, a non-volatile memory, a cache memory, a buffer, a short term memory unit, a long term memory unit, or other suitable memory units or storage units. Memory 120 may be or may include a plurality of, possibly different memory units. Memory 120 may store for example, instructions to carry out a method (e.g. code 125), and/or data such as user responses, interruptions, etc.

[0054] Executable code 125 may be any executable code, e.g., an application, a program, a process, task or script. Executable code 125 may be executed by controller 105 possibly under control of operating system 115. For example, executable code 125 may when executed cause NN training, NN data collection, model creation, early stopping of training, coordination of NN training tasks, NN execution or inference, etc. according to embodiments of the present invention. For the various modules and functions described herein, one or more computing devices 100 or components of computing device 100 may be used. Devices that include components similar or different to those included in computing device 100 may be used, and may be connected to a network and used as a system. One or more processor(s) 105 may be configured to carry out embodiments of the present invention by for example executing software or code. Storage 130 may be or may include, for example, a hard disk drive, a floppy disk drive, a Compact Disk (CD) drive, a CD-Recordable (CD-R) drive, a universal serial bus (USB) device or other suitable removable and/or fixed storage unit. Data such as instructions, code, NN model data, early stopping model data, parameters, etc. may be stored in a storage 130 and may be loaded from storage 130 into a memory 120 where it may be processed by controller 105. In some embodiments, some of the components shown in Fig. 2 may be omitted.

[0055] Input devices 135 may be or may include for example a mouse, a keyboard, a touch screen or pad or any suitable input device. It will be recognized that any suitable number of input devices may be operatively connected to computing device 100 as shown by block 135. Output devices

140 may include one or more displays, speakers and/or any other suitable output devices. It will be recognized that any suitable number of output devices may be operatively connected to computing device 100 as shown by block 140. Any applicable input/output (I/O) devices may be connected to computing device 100, for example, a wired or wireless network interface card (NIC), a modem, printer or facsimile machine, a universal serial bus (USB) device or external hard drive may be included in input devices 135 and/or output devices 140.

[0056] Embodiments of the invention may include one or more article(s) (e.g. memory 120 or storage 130) such as a computer or processor non-transitory readable medium, or a computer or processor non-transitory storage medium, such as for example a memory, a disk drive, or a USB flash memory, encoding, including or storing instructions, e.g., computer-executable instructions, which, when executed by a processor or controller, carry out methods disclosed herein.

[0057] Fig. 3 is a high-level block diagram of an exemplary system which may be used with embodiments of the present invention. Referring to Fig. 3, one or more computer(s) 10 may operate, train, or simulate a NN 12. NN 12 may be represented as data stored on computer 10, e.g. weights, hyperparameters, matrices, etc. NN 12 may be virtual, e.g. having nodes represented by data processed by computer 10. Computer 10 may execute training software 14, including one or more modules or software libraries such as modules or software libraries for controlling or stopping training (e.g. predictive early stopping module 14'), for executing forward and backwards propagation, etc. Predictive early stopping module 14' is in one embodiment a software library included in source code for training software 14. In one arrangement training software 14 may train a NN as is known, and may call predictive early stopping module 14' to determine when to stop training, or to determine whether or not to stop training before training software 14 would otherwise stop training. In one arrangement training software 14 may choose hyperparameters for a NN and test each NN-hyperparameter combination as is known. However, in other embodiments, early stopping functionality, and training, may be performed by different components or modules. While components are shown with respect to one computer 10, such components are not shown for other computers 10 for clarity. Training software 14 may conduct training, during which a NN may be trained using training sets or training data 16, e.g. sample data sets input to a NN during training and used by the NN to alter link weights and produce losses.

[0058] Computer 10 may communicate with learning server or computer 20, which may create, update and/or distribute early stopping module 14' or other software. Communication may be via

network 30, e.g. the Internet. Computer 20 may communicate with computers 10 to receive from training software 14 training information or data 22, and to process and store data 22 in order to update and/or create stopping module 14'. Computer 22 may distribute stopping module 14' or updates to stopping module 14' to computers 10. While in the example shown updates to stopping module 14' are performed by a computer different from a computer using stopping module 14' for training, in other embodiments a computer conducting training may also update or create a stopping module. Other different configurations are possible: for example a computer 10 controlling training and interfacing with a user may outsource some or all of training (including a stopping module 14') to a remote or cloud computer. While stopping module 14' may when executed by a computer 10 indicate when NN training should stop (e.g. by returning True, indicating stop or False, indicating do not stop; other return values indicating stopping or not stopping may be used), in other embodiments the functionality of stopping module may be performed by computer 20 communicating with a computer 10 conducting training, or the functionality of stopping module may be performed by a different remote, e.g. cloud, computer. Examples of functionality stopping module 14' are shown in Tables 1-5 herein.

[0059] Data 22 collected by computer 20 to update stopping module 14' may include for example hyperparameters describing the particular NN and loss data for the NN at the end of each epoch or other period; however other data may be collected. Data 22 may be included (possibly after processing) in models 24 created based on training data collected from training of NNs 12. Models 24 may be for example tree-based regression models. In one embodiment predictive early stopping module 14' may gather data regarding NN learning and transmit the data to learning server 20, so that a model can be updated for use for training other NNs. However, data collection functionality may be performed by different modules.

[0060] Fig. 4A is a flowchart of a method according to embodiments of the present invention. The operations of Fig. 4A may be performed using the systems and modules shown in Figs. 1-3, but may be performed using other equipment. The example operations of Fig. 4A may be used for early stopping when training a NN whose hyperparameters have been chosen, e.g. a "local" mode.

[0061] In operation 400, NN training may be performed on a number of NNs, generating loss data.

[0062] In operation 410, the loss data and other data describing the NNs which generated the loss data may be collected. For example loss data such as time series features, loss values or loss functions, or validation loss functions from training or experiments, may be received by a process

creating training models for use in a stopping procedure. Such data may be processed or normalized, and first- and second-order features may be generated. For example raw time series features may be smoothed by applying a log transformation, followed by minmax scaling.

[0063] In operation 420, one or more models (e.g. models m_μ and m_δ^2 as described elsewhere herein, but other models may be used, such as m_1 , m_2 and m_3 or other models) may be created or built which may take as input data relating to a particular NN and may output predicted loss data and variance data. The model(s) created may include one or more models trained using training losses of a plurality of NNs including NNs other than a NN which will eventually use (or have used on the NN) the model to determine an efficient time to stop training. (However, training for a target NN may be included in an early stopping model used for that particular NN.) E.g. in some embodiments, training data and results, including data describing a NN, may be collected for many, e.g. thousands of NN during or after training of those NNs, and training a NN may use models created based on these other NNs to determine a best or likely best time to stop training the NN. In one embodiment a model may include tree based regression models. The tree-based models may, when presented with input data such as a particular NN's loss data and hyperparameters, provide estimates of the mean or average expected change in the loss curve, and a variance. In other embodiments, models may return upper and lower bounds of a range, for example quantile losses that estimates a lower percentile and an upper percentile, e.g. the 5th and 95th quantile change, in the loss curve. Other or different models may be used; for example the lower percentile model need not be used.

[0064] In operation 430, training may begin or continue (e.g. for a next training interval or epoch) on a NN, e.g. by a system such as shown in Fig. 3, where training data is presented to a NN over epochs. Loss data may be computed and/or returned periodically, for example after each forward and backward pass, or in or after each epoch. In some embodiments, the loss data for a NN being trained using a set of models determining a likely best stopping time may be used to update those same models. Actual loss data may be generated, for example the most recent loss data generated or computed, e.g. the loss data for the most recent or latest training epoch completed, or another measure of actual loss data for the NN for a specific period, not including past or historic loss data.

[0065] Typically, the use of an early stopping prediction may be delayed near the beginning of training to prevent early stopping if stopping conditions are met in an early iteration or epoch (e.g. epoch 5). Stopping conditions may include a probability being compared to a threshold, the value

of a wait counter, actual loss is less than predicted loss, etc. For example, a process may wait for an initial number of iterations or epochs, e.g. 10 epochs, or another number of epochs or another delay period, for a target NN being trained, before determining if training should be stopped. The epoch number may be inferred, e.g. from the number of items in the ordered series of loss data.

[0066] In operation 440, a probability of improvement over a best loss seen for this NN may be calculated. For example, a model may be used to determine a mean expected best loss for the NN, and another model may be used to determine the variance of that mean, and calculations (e.g. as shown in the example in Table 3) may be used to determine the probability.

[0067] The probability of a mean or predicted best loss (which may be an average predicted best loss over a set of NNs sharing similar characteristics) improving over a best loss seen so far, e.g. `best_metric` (which may be the best loss so far for a particular NN, or a group of NNs having hyperparameters chosen), may be calculated by determining the variance for the mean, calculating the square root of the variance, and using a function such as the known cumulative distribution function (CDF) having as inputs the mean, the square root of the variance, and the best loss seen so far. Other methods of calculating the probability of improvement may be used.

[0068] For example, an expected training loss may be determining using data such as a set of model parameters, such as hyperparameters for the NN and loss history for the NN, and the current loss for the NN (loss history may in some embodiments include the current loss). A loss history may be a vector or ordered series of losses each recorded at a specific point in time, for example at the end of an epoch. For example, a loss history may be an ordered series $(y_0, y_1, \dots, y_{t-1})$, where y_x is the loss history at time x , e.g. x being an epoch, or an iteration. In one embodiment y_x is the last loss for the last batch for epoch x , but other methods may be used. The loss history input to the model may be the same format (e.g. same type of time history created, same value range) as the loss history used in the model and collected from other NNs. The current loss may be for example the last loss recorded in the last forward/backward pass. Other NN parameters may be used. Such parameters may be input to the model created in operation 420, e.g. a model which has been built using data (e.g. loss data, hyperparameters) from many other NNs, even NNs with very different structures and functions from the NN currently being evaluated. Typically, loss series data is processed, e.g. normalized where the normalized data has parameters extracted (e.g. derivatives), and this processed data is used by the model as characterizing the NN.

[0069] If a model is used having a different model or sub-model for each epoch, a process and model may infer the epoch or time period for the training of the target NN from the number of time series data items, e.g. if there are 10 entries in an ordered series of loss data, it may be inferred there have been 10 epochs of training. Each model progressing along an increasing number of epochs or other periods may use all the information seen so far, e.g. a model for epoch 10 may use 10 epochs of data, while a model for epoch 100 may use 100 epochs of data.

[0070] In operation 450, it may be determined if the actual training loss is less than a historic minimum training loss.

[0071] In operation 460, if the current or actual training loss is less than a historic minimum training loss the wait value is not increased, and may be reset or set to zero, or to a value indicating waiting should begin at the beginning or for the maximum of the wait period. In some embodiments, an integer wait value is used to determine a period of waiting or “patience” where if no improvement is seen over the period, training is stopped. The period may be measured in training intervals or epochs, rather than absolute time, but in other embodiments other periods may be used. That the current or actual training loss is less than a historic minimum training loss may indicate that a new “high” in training success has been reached, so a “patience” period may be reset.

[0072] In operation 470, if the actual training loss is not less than a historic minimum training loss, a wait value may be incremented or increased, e.g. by a convenient integer count value such as 1. Typically, if the loss function of a NN is decreasing it indicates an improvement of the functionality of the NN during training. A historic minimum training loss may be for example the lowest training loss value in the loss history for the NN. A wait value may be for example an integer, initialized to a value such as zero, which is incremented by one. Such an increment or increase may take place if training loss does not improve, for example over a most recent measurement, a determination which may be made if, at a point in time, actual or current training loss is not less than a historic minimum training loss. Another measurement of no improvement may be used.

[0073] As with other values, thresholds and comparisons discussed herein, the “or equal” in the “greater than” and “less than” decisions may be swapped: for example in some embodiments if the actual training loss is less than or equal to a historic minimum training loss the wait value is

not increased, and if the actual training loss is not less or equal to a historic minimum training loss the wait value is increased.

[0074] In operation 480, if the probability of improvement over a best loss seen for this NN is less than or equal to a probability threshold, training may be stopped. If the probability of improvement over a best loss seen for this NN is not less than or equal to a threshold, the process may continue to operation 490.

[0075] In operation 490, if the wait value is greater than or equal to a “patience” value or threshold, training may be stopped; and if the wait value is not greater than or equal to a “patience” value or threshold, training may be continued at operation 430. Thus if neither of the “stopping conditions” of operations 480 and 490 are true, training may continue at operation 430. E.g. if the wait value is not greater than or equal to a threshold and if the probability of improvement is not less than or equal to a probability threshold, training may continue at operation 430.

[0076] Training may be stopped by a process, such as a predictive early stopping module (e.g. performing operations such as shown in Tables 1-5). Such a process may return a stop/True value or a continue training/False value, or signal training should stop or continue by another process. While in some embodiments a module which determines whether or not to stop training is described as a module separate from a module or system which conducts training (e.g. executes forward/backward passes, adjust weights, etc.) in other embodiments a stopping decision may be made by a training module itself.

[0077] Other or different operations may be performed.

[0078] Fig. 4B is a flowchart of a method according to embodiments of the present invention. The operations of Fig. 4B may be performed using the systems and modules shown in Figs. 1-3, but may be performed using other equipment. The example operations of Fig. 4B may be used for early stopping when searching for a set of hyperparameters for a NN. In general, the operations of Fig. 4B may be similar to those in Fig. 4A, but for certain differences described below.

[0079] In operation 500, NN training may be performed on a number of NNs, generating loss data.

[0080] In operation 510, the loss data and other data describing the NNs which generated the loss data may be collected and used in operation 520 to create models.

[0081] In operation 525, a set of hyperparameters may be chosen for a NN using known methods, e.g. traversing or randomly choosing an entry in a grid, each grid entry including a set of hyperparameters.

[0082] In operation 530, training may begin or continue (e.g. for a next training interval or epoch) on a NN having hyperparameters chosen in operation 525. The use of an early stopping prediction may be delayed near the beginning of training. The training of a specific NN having a specific or unique set of hyperparameters chosen in operation 525 may be called an “experiment”, and a number of experiments may take place. The resulting best loss for each NN in an experiment may be recorded. In another embodiment one best loss value over all different NN-hyperparameter combinations seen thus far may be recorded (e.g. as `best_metric`); when a new best loss occurs, it may replace the best loss value.

[0083] In operation 540, if the probability of improvement over a best loss seen for other NNs is less than or equal to a probability threshold, a wait value may be increased. Such an increase is typically by an integer value of 1.

[0084] In operation 550, if a wait value is greater than or equal to a patience value or threshold, then training this particular NN may be stopped (in one embodiment this is effected by a function returning “true”), and the process may proceed to operation 570.

[0085] In operation 560, training may continue on the same, current, NN (in one embodiment this is effected by a function returning “false”). If the training for this NN has completed (e.g. because of a “local” early stopping determination such as that made by a process such as in Fig. 4A, or another determination or measure), the process may continue to operation 570; otherwise the process may continue training this NN by continuing to operation 530.

[0086] In operation 570, if there are more hyperparameter sets to consider, or that a process is set to consider, a process may continue at operation 525 for a new NN; otherwise the process may stop. In some embodiments, a certain number of hyperparameters may be set as a limit to test; in other embodiments other measures of “more hyperparameter sets to consider” may be used. In some embodiments not all hyperparameter sets are considered.

[0087] Other or different operations may be performed.

[0088] The expected training loss may take multiple forms. For example, a mean or average expected change in the loss curve for other NNs fitting the same category or description (e.g. categorized in the same leaf on a tree-based regression model) may be or be part of an expected training loss. One tree model that may be used is the Light GBM model, a gradient boosting framework that uses a tree based learning algorithm. Thus determining using parameters training loss an expected loss may include obtaining from a leaf of at least one tree data structure the

expected training loss. Expected loss may also include a range of possible losses or a minimum and/or a maximum loss for a certain percentage of other NNs fitting the same category or description, for example a lower percentile and an upper percentile, e.g. the 5th and 95th quantile change, in the loss curve. In one embodiment such a range is determined from two different models, each providing an endpoint for the range. Determining an expected training loss may include determining the minimum of: a mean expected training loss (possibly less the loss threshold as one manner of adjustment); and an estimate of a training loss occurring in a certain percentile of the set of cases. For example, the returned value may be the lower of a mean expected loss, and a maximum of a range, e.g. the loss where 95% of the NNs in the model matching the input to the model (e.g. in the leaf of the target NN or having hyperparameters similar to the target NN) have a loss greater than or equal to the actual current loss.

[0089] Figs. 5A and 5B depict example prediction loss curves, according to one embodiment of the invention. Each of Figs. 5A and 5B depict a set of predictions made by an example set of models for a different set of NN instances. As can be seen in Fig. 5, embodiments may accurately estimate the minimum convergence values for the validation loss hundreds of training cycles, periods or epochs in advance of when a minimum actually occurs. In Fig. 5, dots 600 depict a predicted final loss (not loss at the epoch shown in the X axis), for a specific category of instances of NN described by certain hyperparameters (typically a range of hyperparameters) and loss history (typically a range of loss history data) over a series of epochs, shown in the X-axis. The predicted loss may be a mean of loss for all NNs within a certain category of instances, possibly normalized when displayed in Fig. 5 to the range of loss values of the NN whose actual loss over epochs is shown in orange lines 602. The loss data for each NN in a leaf is typically stored as normalized data, is typically unitless, and may be converted back to a range appropriate for the particular NN being analyzed, for example by the model returning the loss. The representation or data for each leaf in a model may include loss value(s) normalized to a standard range, e.g. 0-1 or another range, and this may be converted back to a range appropriate for the particular NN being analyzed. The Y axis indicates loss over epochs shown in the X axis. The lower bound of blue lines 604 for each epoch shows the lower (e.g. 5th) percentile change, or lower percentile loss, where below this level 5% of instances NNs have losses. The upper bound of lines 604 for each epoch shows the upper (e.g. 95th) percentile change, or upper percentile loss. It can be seen that actual loss 602 for a NN converges with predicted final loss 600 as the number of epochs increase.

In Fig. 5B, a stopping algorithm according to embodiments of the present invention has been applied at point 606, to prevent future training where it is predicted that no or insignificant loss improvement will take place.

[0090] Figs. 6A and 6B depict actual changes in loss values compared with predicted changes in the loss curves, according to one embodiment of the invention. Fig. 6A shows the comparison at epoch 100, for a sample set of data, and Fig. 6B shows the comparison at epoch 500. The Y axis represents actual loss data, and the X axis represents predicted loss. As can be seen, the example model used tends to make better predictions at epoch 500 because more of the learning curve has been used as inputs. Color key 650 shows a likelihood score representing prediction confidence for each entry in the graph, where a higher number is a higher confidence regarding a prediction.

[0091] In Figs. 5 and 6 it can be seen that some embodiments may accurately estimate the minimum convergence values for the validation loss hundreds of epochs in advance in some cases.

[0092] Parameters or settings to control early stopping may be predefined or provided by a user. A probability threshold and a patience or wait threshold, as used in Tables 1-5 herein, may be received or predefined. A threshold such as an accuracy threshold may be predefined or received from a user, which may define or determine how close an actual current training result (e.g. loss) needs to be to the predicted target we to stop learning. Threshold may be expressed for example as a percentage, but other ways of expressing a threshold may be used. For example a threshold may be within a range of 0-1.0, such that if the threshold value is 0.01, a training process will be stopped if the actual or current NN loss value is within 1% of the anticipated change.

[0093] Figs. 7A and 7B depict example early stopping results, according to one embodiment of the invention with a particular example set of input data. A test was conducted for an example embodiment stopping algorithm by first training a set of models to a desired loss value using conventional stopping mechanisms – e.g. stopping when loss reaches a plateau. These models were then used with different parameters for a stopping process. Time saved was measured as the difference between the time taken to run the algorithm (and train each NN) to the desired validation loss without the use of any stopping, and the time taken to get to within a threshold value of the predicted validation loss while training the NNs, using an embodiment of the present invention. Error is measured as the difference between the predicted change in the loss value, and the actual change in the loss value. If the embodiment works well, the error may be approximately the same as the threshold value parameter.

[0094] The graph in Fig. 7A shows the results of early stopping algorithm with various threshold and patience parameter values. In Fig. 7A the Y axis shows the time saved (the final values shown in boxes 806 being for patience value 100) by certain stopping policy (e.g. a set of input control values such as patience and accuracy threshold) in percentage terms (where percentage is converted to a 0-1 scale; 0.2 means 20% reduction in training time). The X axis is the value of a patience input control parameter. Line 800, shown in orange, shows time saved for an input accuracy threshold value of 0.01, line 802, shown in blue, shows time saved for an input accuracy threshold value of 0.05, and line 804, shown in green, shows time saved for an input accuracy threshold value of 0.1. An accuracy threshold may be expressed in a percentage such that a value of 0.01 means 1%.

[0095] The graph in Fig. 7B shows the results of early stopping algorithm with various threshold and patience values. In Fig. 7B the Y axis is the difference or error between the change in loss achieved during full training (e.g. without early stopping, rather the when training algorithm stops naturally according to prior art methods) and the change in loss at the predicted stop point. In Fig. 7B, if the actual change to minimum using prior art methods was 97% and an early stopping algorithm stopped training at 96%, the error will be 1% or 0.01. The X axis is the value of the patience input control parameter. Line 810 (orange) shows the error or difference input accuracy threshold value of 0.01, line 812 (blue) shows error or difference for an input accuracy threshold value of 0.05, and line 814 (green) shows error or difference for an input accuracy threshold value of 0.1. Fig. 7B shows that using one embodiment, error may follow the threshold value quite closely.

[0096] For example, at a threshold value of 0.01, and patience value of 100, an embodiment can reduce training time by an average of 20% with a 4% error.

[0097] Embodiments may provide early stopping in a “global” mode to choose hyperparameters for a NN, and/or in a “local” mode with a single experiment, e.g. when hyperparameters have been chosen and training of a specific model is taking place. An embodiment may cause stopping of training in a “global” mode as soon as a probability that the likely best loss of the NN is lower than the best loss of other NNs which have been trained with other hyperparameters is lower than a threshold, or when a wait or patience value exceeds a threshold. Thus the operations shown in Fig. 4A in one embodiment may cause early stopping of NN training for a NN of a known configuration, and the embodiment shown in Fig. 4B may allow choosing among the best NN

structure. For example, a number of different NN structures (e.g. having different number of neurons, different layers, different types of layers, etc.) may be trained, and early stopping may be applied to NNs that do not show improvement after a configurable amount of time. Training then may take place on a chosen NN configuration, with or without early stopping.

[0098] In a global mode hyperparameters for a NN may be chosen as known in the art, e.g. using a grid where each cell indicates a set of hyperparameters and choosing cells (e.g. sets of hyperparameters) randomly or in an ordered manner. For each NN with a set of hyperparameters, training may take place, and an early stopping method may be applied which may stop training if it is predicted, or it is likely, that the best loss achievable for the current NN is not better (e.g. less than) the best loss achievable for the NNs already trained (with other hyperparameters). Thus training a NN with “worse” hyperparameters may be avoided. A “sweep”, “start” or delay variable may cause the early stopping algorithm to not be used or to not cause early stopping before a certain number of experiments on NNs with various hyperparameters to be completed, e.g. the number of initial experiments to wait to complete in a hyperparameter sweep, before starting to make predictions.

[0099] An embodiment may cause stopping of training in a “local” mode as soon as a probability of improvement in the loss of the NN is less than a threshold, or a wait value is greater than a wait threshold. An alternate embodiment may cause stopping of training as soon as a minimum predicted validation loss is achieved (possibly within some tolerance or accuracy) or when a wait or patience value exceeds a threshold.

[00100] Two types of stopping rules may be used (e.g. the global and local embodiments as described), and input or received parameters (e.g. user parameters) may specify which rule to use.

[00101] One embodiment can be expressed as using a combination of modules or algorithms, described below in pseudocode. Table 1 shows an example algorithm for causing early stopping during training of a NN in a “local” mode: e.g. early stopping for a NN whose hyperparameters have been chosen. Table 2 shows an example algorithm for early stopping in a “global” mode, e.g. to determine if to stop training if the probability of loss improvement when compared with other NNs with different hyperparameters is less than a threshold. Example procedures in Tables 1 and 2 may be called by a NN training or hyperparameter determination process to determine when to stop training a particular NN. Table 3 shows an example algorithm for generating a probability that the best or ultimate loss for a NN will be less than a “best metric”,

e.g. the best actual loss seen when using the NN in local mode or the best actual loss seen when using another NN with other hyperparameters in global mode. The example procedure in Table 3 may be called by the example procedures in Tables 1 and 2. Other algorithms of code may be used, and while the various Tables show functionality split into different parts or algorithms, in other embodiments functionality may be combined in one unit or organized in different units.

[00102] As can be seen in Table 3, to calculate a probability in one embodiment, it is assumed that the distribution around the predicted value is normal or Gaussian. A sample minimum value may be compared to this to obtain the probability that a better minimum value will be seen. In one embodiment a model may return a mean and variance, and this may be used to calculate the probability. A variance may be returned that may be used to create an interval within which most occurrences (e.g. 90% of the recorded occurrence) may fall.

[00103] A “threshold” parameter may define a threshold for a probability of improvement of the current loss curve, where if the probability of improvement is below the threshold stopping may occur. Other thresholds may be used, e.g. as in Tables 4 and 5, where a threshold may be a loss amount or range, where if the actual loss is within a threshold of predicted loss, stopping may occur. The threshold parameter may be used, for example, in:

- Local Mode: Training is stopped when the predicted probability of improvement is lower than the threshold.
- Global Mode: A counter (e.g. to be compared to a patience parameter) is incremented every time the predicted probability of improvement is less than the threshold.

[00104] A “patience” parameter may determine or define how many epochs or other units a process will wait for the validation loss of a particular NN model being trained to fall within a model’s prediction, prediction interval or value. In some embodiments, in a “local” mode, evaluating a certain NN, patience may define the number of evaluations to wait without seeing any improvement in the loss curve, and may reset every time there is an improvement in the loss curve. If the actual or current NN validation loss is within the predicted interval at a certain period of time, a patience value or counter may reset whenever there is improvement over time in the loss. If the loss fails to improve, the patience value or parameter may be incremented, and training the model may be stopped once the patience value or parameter exceeds the patience parameter.

[00105] In some embodiments, in a “global” mode, evaluating a certain NN, patience may define the number of evaluations or epochs to wait where the probability of the loss curve

improvement over the best seen so far with a NN having different hyperparameters is less than a threshold. In a 'local' mode, patience may define the number of evaluations where the predicted probability of improvement is less than the threshold.

[00106] Tables 1 and 2 are shown below:

Algorithm 1 Early Stopping (Local)

Global Variables:

wait, initialized to 0 // wait is a wait parameter to be incremented if no improvement

Input:

loss_history = $[y_0, y_1, \dots, y_{t-1}]$ // Historical loss values for the NN to be analyzed for stopping training. Typically each y is collected at the end of each sequential epoch or other period of time, but other ways of organizing a series of loss data may be used. Data is for epochs or periods labelled starting at zero. In one embodiment normalization and conditioning are performed in the model used in PREDICT, shown in Table 3. loss_history excludes the current loss value y_t

current = y_t // the current loss value for the NN, e.g. value subsequent to y_{t-1}

Φ = hyperparameters for the NN

patience = $Z > 0$ // threshold for wait

threshold = $R \in [0, 1]$ // threshold for probability, range 0-1

Output:

True/False // True tells calling procedure to stop training; False=keep training

```

1: procedure EARLY STOPPING (loss_history, current,  $\Phi$ , threshold, patience) // call
   procedure for local stopping with input
2: current_min = min(loss_history) // current_min is the lowest value in loss_history for NN
3: current_best = min(current, current_min) // minimum from loss history and current loss
4: prediction, interval, probi = Predict(loss_history,  $\Phi$ , current_best) // Example of predict is
   described in Table 3; interval and prediction may not be used, but may be used in some
   embodiments
5: if current < current_min then // If there is improvement (current NN loss better than historic),
   reset the wait value
6:   wait = 0
7: else // If no improvement is seen, increment the wait value
8:   wait = wait + 1
9: if probi <= threshold then // If probability of improvement less than threshold, stop training;
10:   return True // True is a determination to stop training; signals the procedure calling
   EARLY STOPPING to stop training the NN
11: if wait >= patience then // If the wait value is greater than or equal to patience value there has
   been no improvement over some time, so stop training
12:   return True

```

13: return False // signal to calling procedure to continue training NN; if any True condition is not met, return False; False is a determination to continue training;

Table 1

Algorithm 2 Early Stopping (Global)

Global Variables:

wait = 0

Input:

current_loss_history = $[y_0, y_1, \dots, y_{t-1}]$

best_metric = $\min(\min(\text{loss_history}_0); \min(\text{loss_history}_1); \dots \min(\text{loss_history}_n))$ //

loss_history_x is the loss history from NN_x trained NNs having hyperparameters different from NN being trained currently; best_metric is lowest loss found among loss histories for all other NNs; alternately one “best seen so far” variable may be used

patience = $Z > 0$

Φ = hyperparameters for the NN

threshold = $R \in [0, 1]$ // threshold for probability, range 0-1

Output:

True/False // True tells calling procedure to stop training; False=keep training

1: procedure EARLY STOPPING(loss_history, current_loss_history, Φ , start, interval, threshold, patience)

2: prediction, interval, prob_i = Predict(loss_history, Φ , best_metric)

3: if prob_i ≤ threshold then // If the probability this NN will be trained to a loss better than best seen with other NNs is less than a probability threshold, increase wait else if probability > threshold, return false below and continue training

4: wait = wait + 1

5: if wait ≥ patience then

6: return True // Stop early; no probability seen over a certain period that this NN will have better loss than others, thus stop training this NN; True is a determination to stop training

7: return False // If the probability NN will be trained to a loss better than best seen with other NNs is greater than threshold, do not stop training; False is a determination not to stop training

Table 2

[00107] The example procedures shown in Tables 1 and 2 call or execute an example procedure such as shown in Table 3, to return a probability for improvement. The example

procedures shown in Tables 1 and 2 may in turn be called or executed by a procedure choosing hyperparameters for a NN, or a procedure training a NN for which hyperparameters have been chosen. The functionality shown across the Tables in the present application may be organized differently, and for example do not need to be divided across modules as disclosed herein. Models such as m_μ and m_δ^2 used in the algorithm in Table 3, or other models such as those used in Table 5, may accept a loss history and convert that history to features or other derived values, such as by using preprocessing and methods such as extracting features. Thus in some embodiments, a model may include code, or be in part an executable module, deriving the relevant features or values from a loss sequence.

Algorithm 3 Predict

Input:

current_loss_history = $[y_0, y_1, \dots, y_{t-1}]$ // For NN being evaluated
 Φ = hyperparameters
best_metric = $R > 0$ // if “local” may be best loss seen for specific NN; if global may be best loss across other NNs trained, typically with different hyperparameters

Output:

prediction = $R > 0$ // predicted loss history, expressed as mean below
interval = $[R > 0; R > 0]$
 $\text{prob}_i = R \in [0, 1]$ // likelihood, between 0 and 1, that this NN’s loss will ever be better than input best_metric

- 1: procedure PREDICT(loss_history, Φ , best_metric)
- 2: origin = y_0 // origin, used to normalize or re-scale, is set to the first loss recorded for the NN
- 3: current_min = $\min(\text{loss_history})$
- 4: mean = origin $\times (1.0 - m_\mu(\text{loss_history}, \Phi))$ // input loss_history and hyperparameters to model to receive from model mean, the un-normalized likely best loss predicted to be achieved for this NN
- 5: var = origin $\times (m_\delta^2(\text{loss_history}, \Phi))$ // Input data to model to receive from model predicted variance
- 6: std = square root of var
- 7: max = mean + $1.96 \times \text{std}$
- 8: min = mean - $1.96 \times \text{std}$
- 9: $\text{prob}_i = \Phi_{\text{mean,std}}(\text{best_metric})$ // Probability of improvement from a normal CDF (cumulative distribution function) parameterized by the mean and standard deviation; $\Phi_{\text{mean,std}}$ is CDF, which expresses a probability distribution
- 10: return mean, [min, max], prob_i // [min, max] is pair of values returned as interval

Table 3

[00108] Tables 4 and 5 depict an alternate embodiment of a set of procedures for early stopping. The example procedure shown in Table 4 may in turn be called or executed by a procedure training a NN for which hyperparameters have been chosen. Tables 4 and 5 are as follows:

Global Variables:

wait, initialized to 0 // wait is a wait parameter to be incremented if no improvement

Input:

loss_history = $[y_0, y_1, \dots, y_{t-1}]$ // historical loss values for the NN to be analyzed for stopping training typically at the end of each sequential epoch or other period of time, but other ways of organizing a series of loss data may be used; for epochs or periods labelled starting at zero

current = y_t // the current loss value for the NN, e.g. value subsequent to y_{t-1}

Φ = hyperparameters for the NN

patience = $Z > 0$ // threshold for wait

threshold = $R \in [0, 1]$

Output:

True/False

```

1: procedure EARLY STOPPING (loss_history, current,  $\Phi$ , threshold, patience) // call
   procedure with input
2: current_min = min(loss_history) // current_min is the lowest value in loss_history for NN
3: current_best = min(current, current_min)
4: prediction, interval = Predict(loss_history,  $\Phi$ , threshold) // Predict described in Table 5;
   interval may not be used, but may be used in some embodiments, so may be returned
5: if current < current_min then // If there is improvement (current NN loss better than historic),
   reset the wait value
6:   wait = 0
7: else // If no improvement is seen, increment the wait value
8:   wait = wait + 1
9: if current_best <= prediction then // If the minimum is less than or equal to the predicted
   value, stop training; prediction is adjusted in Predict line 4 by threshold
10:   return True // True is a signal to the procedure calling EARLY STOPPING to stop
   training the NN
11: if wait >= patience then // If the wait value is greater than or equal to patience value, stop
   training
12:   return True // Stop training

```

13: return False // signal to calling procedure to continue training NN; if any True conditions not met, return false

Table 4

Parameters:

Origin = y_0 // origin, used to normalize or re-scale, is set to the first loss recorded for the NN

Input:

loss_history = $[y_0, y_1, \dots, y_{t-1}]$

Φ = hyperparameters for the NN

threshold = $R \in [0, 1]$ // range of 0-1

Output:

Prediction = $R > 0$

interval = $[R > 0, R > 0]$ // A range bounded by the highest and lowest X%, e.g. 5% of returned values

1: procedure PREDICT (loss_history, Φ , threshold) // Algorithm called by EARLY STOPPING to return loss estimate and range

2: origin = y_0

3: mean = origin $\times (1.0 - m_1^{t-1}(\text{loss_history}, \Phi) - \text{threshold})$ // Adjust mean prediction based on provided threshold parameter; m_1 is model returning an estimated mean; note this example embodiment of a “Predict” operation performs calculations in percentage then converts to an actual loss by multiplying by origin; mean, min and max are on scale of 0-1

4: min = origin $\times (1.0 - m_2^{t-1}(\text{loss_history}, \Phi))$ // m_2 is model returning an upper or high bound for an estimated range

5: max = origin $\times (1.0 - m_3^{t-1}(\text{loss_history}, \Phi))$ // m_3 is model returning a lower bound for an estimated range

6: return min(mean,max), [min, max] // Ensures that the adjustment to the mean value does not exceed the upper limit of the interval

Table 5

[00109] In the two algorithms in Tables 4 and 5, used in a “local” mode, EARLY STOPPING (Table 4) may be used by a training process to determine when to stop training; a return value of true means stop training, and false means continue training. EARLY STOPPING may take as inputs the loss history and current loss for a NN, the hyperparameters for the NN, and parameters such as patience and a threshold. EARLY STOPPING may call Predict (Table 5) which may use the sets of models, e.g. m_1 , m_2 , and m_3 discussed elsewhere. A model may determine the number of loss values in a loss history and determine from this which epoch number

to use and thus which specific model to use (e.g. m_1^X , m_2^X and m_3^X for epoch X). Input to a model or to a model process may determine which specific model to use: for example if input to a model describes X epochs, the model(s) for epoch X will be used (or epoch X-1 if numbering starting at 0 is used). A model may process loss history data (e.g. normalization, calculation of derivatives) before use. The output from a model may be created by finding the relevant leaf into which the NN is categorized and returning the average of the loss values for all NNs in that relevant leaf. For example, if m_1 is a model returning an estimated mean, m_1 may pre-process an input loss history (e.g. normalize), determine hyperparameters for the pre-processed data (e.g. derivatives, curve analysis), find the leaf matching or including the range in which the hyperparameters exist, and return the mean of loss values at the relevant epoch for NNs in that leaf. Procedure PREDICT in Table 5 may return loss values or ranges which are normalized to the particular NN being analyzed.

[00110] Internal to Predict, the loss returned by the models m_x may be expressed in percentages (e.g. 0-1) and normalized to loss values relevant to a particular NN by multiplying by the first loss value seen for that NN (e.g. “origin”). The use of the threshold in Predict modifies the mean prediction so that the threshold (e.g. the level of accuracy set as a parameter, predefined or set by a user) is taken into account when the mean is compared in EARLY STOPPING. For example, the threshold value may “push” the mean value. For example, if loss starts at 100, and a 97% improvement is expected based on a model, if a user sets a threshold at 5%, then stopping may occur at 92% loss improvement.

[00111] The algorithms in Table 4 and Table 5 maintain a range bounded by “max” or the value of m_3 . However, in the embodiment shown in Table 4 and Table 5 the range, and min, are not used. In other embodiments, such a range, or min, may be used. For example, a returned min or range may be used to ensure a user’s confidence in the process, or may be provided on a user interface.

[00112] In one embodiment, the loss values in a model are stored as a percentage, or a range from 0-1, and then may be converted to a real loss value when returned from the model. For example, this conversion may take place by multiplying the returned percentage by “origin”, the first loss value received for a NN.

[00113] The input to the stopping process may include model parameters describing a particular NN at a particular point in time in training for that NN, e.g. during or after a particular

epoch in training. Model parameters may include, for example, the loss history, a current loss for the NN (typically measured at the end of an epoch, but a different way of measuring current loss for a time-series item may be used), hyperparameters for or some other description of the NN, and possibly user variables such as patience or a threshold for waiting, and an accuracy threshold defining a range within which if the expected loss falls training can be stopped. In some embodiments a “loss history” may include a current loss, and thus a current loss need not be a separate data item.

[00114] Embodiments of the present invention may improve prior NN training by making training or learning more efficient, by stopping training at a point where little or no further improvement is expected, or less than a certain amount of improvement is expected. Embodiments of the present invention may improve and make more efficient prior NN hyperparameter selection by stopping training of a NN with a set of hyperparameters at a point where it is unlikely that the current NN can be trained to have a loss better than the best seen so far for other NNs. This may be compared with prior methods, which notice improvement training after-the fact: embodiments of the present invention may use past data to predict when such stagnation may occur, or to predict a best loss to be found in the future: for example embodiments of the invention may predict future improvement, or the odds of improving in the future. This may save resources such as compute or processor time. Embodiments may provide an improvement over prior art learning mechanisms in that specific aspects of a NN such as architecture and application are not considered, as certain embodiments may be applied across a range of NN types, applications and architectures. Embodiments of the present invention may deliver better predictions by returning (using an internal model) a predicted minimum, e.g. the best a NN is predicted to achieve, as opposed to a specific predicted value for a predicted specific time.

[00115] Embodiments were tested in a “global mode” of stopping while choosing hyperparameters by optimizing a 6-layer Convolutional Neural Network on the CIFAR10 dataset (a known dataset including thousands of color images in different classes), using an SMAC (sequential model-based algorithm configuration) optimizer. The SMAC optimizer was executed with and without an embodiment of an early stopping algorithm described herein. The total number of epochs required to achieve a comparable test set error was compared, and in one example a comparable error on the test set of CIFAR10 was achieved, using early stopping, in 300 fewer epochs than without.

[00116] One skilled in the art will realize the invention may be embodied in other specific forms without departing from the spirit or essential characteristics thereof. The foregoing embodiments are therefore to be considered in all respects illustrative rather than limiting of the invention described herein. Scope of the invention is thus indicated by the appended claims, rather than by the foregoing description, and all changes that come within the meaning and range of equivalency of the claims are therefore intended to be embraced therein.

[00117] In the foregoing detailed description, numerous specific details are set forth in order to provide an understanding of the invention. However, it will be understood by those skilled in the art that the invention can be practiced without these specific details. In other instances, well-known methods, procedures, and components, modules, units and/or circuits have not been described in detail so as not to obscure the invention. Some features or elements described with respect to one embodiment can be combined with features or elements described with respect to other embodiments.

[00118] Although embodiments of the invention are not limited in this regard, discussions utilizing terms such as, for example, “processing,” “computing,” “calculating,” “determining,” “establishing,” “analyzing,” “checking”, or the like, can refer to operation(s) and/or process(es) of a computer, a computing platform, a computing system, or other electronic computing device, that manipulates and/or transforms data represented as physical (e.g., electronic) quantities within the computer’s registers and/or memories into other data similarly represented as physical quantities within the computer’s registers and/or memories or other information non-transitory storage medium that can store instructions to perform operations and/or processes.

[00119] Although embodiments of the invention are not limited in this regard, the terms “plurality” and “a plurality” as used herein can include, for example, “multiple” or “two or more”. The terms “plurality” or “a plurality” can be used throughout the specification to describe two or more components, devices, elements, units, parameters, or the like. The term set when used herein can include one or more items. Unless explicitly stated, the method embodiments described herein are not constrained to a particular order or sequence. Additionally, some of the described method embodiments or elements thereof can occur or be performed simultaneously, at the same point in time, or concurrently.

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76		Attorney Docket Number	P-587424-US
		Application Number	Not yet assigned
Title of Invention	SYSTEMS AND METHODS FOR PREDICTIVE EARLY STOPPING IN NEURAL NETWORK TRAINING		
The application data sheet is part of the provisional or nonprovisional application for which it is being submitted. The following form contains the bibliographic data arranged in a format specified by the United States Patent and Trademark Office as outlined in 37 CFR 1.76. This document may be completed electronically and submitted to the Office in electronic format using the Electronic Filing System (EFS) or the document may be printed and included in a paper filed application.			

Secrecy Order 37 CFR 5.2

☐	Portions or all of the application associated with this Application Data Sheet may fall under a Secrecy Order pursuant to 37 CFR 5.2 (Paper filers only. Applications that fall under Secrecy Order may not be filed electronically.)
--------------------------	---

Inventor Information:

Inventor 1					Remove
Legal Name					
Prefix	Given Name	Middle Name	Family Name	Suffix	
	Dhruv		NAIR		
Residence Information (Select One) ☒ US Residency ☐ Non US Residency ☐ Active US Military Service					
City	New York	State/Province	NY	Country of Residence	USA
Mailing Address of Inventor:					
Address 1	240 Richardson St				
Address 2	Apt 3A				
City	New York	State/Province	NY		
Postal Code	11222	Country	USA		
All Inventors Must Be Listed - Additional Inventor Information blocks may be generated within this form by selecting the Add button. Add					

Inventor 2					Remove
Legal Name					
Prefix	Given Name	Middle Name	Family Name	Suffix	
	Gideon		MENDELS		
Residence Information (Select One) ☒ US Residency ☐ Non US Residency ☐ Active US Military Service					
City	New York	State/Province	NY	Country of Residence	USA
Mailing Address of Inventor:					
Address 1	1 University Place 2B				
Address 2					
City	New York	State/Province	NY		
Postal Code	10003	Country	USA		
All Inventors Must Be Listed - Additional Inventor Information blocks may be generated within this form by selecting the Add button. Add					

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76		Attorney Docket Number	P-587424-US
		Application Number	Not yet assigned
Title of Invention	SYSTEMS AND METHODS FOR PREDICTIVE EARLY STOPPING IN NEURAL NETWORK TRAINING		

Inventor 3 				
Legal Name				
Prefix	Given Name	Middle Name	Family Name	Suffix
	Nimrod		LAHAV	
Residence Information (Select One) ☒ US Residency ☐ Non US Residency ☐ Active US Military Service				
City	Hoboken	State/Province	NJ	Country of Residence USA
Mailing Address of Inventor:				
Address 1	333 River Street			
Address 2	Apt 508			
City	Hoboken	State/Province	NJ	
Postal Code	07030	Country	USA	
All Inventors Must Be Listed - Additional Inventor Information blocks may be generated within this form by selecting the Add button. 				

Correspondence Information:

Enter either Customer Number or complete the Correspondence Information section below. For further information see 37 CFR 1.33(a).	
☐ An Address is being provided for the correspondence information of this application.	
Customer Number	49443
Email Address	USPTO@PearlCohen.com  

Application Information:

Title of the Invention	SYSTEMS AND METHODS FOR PREDICTIVE EARLY STOPPING IN NEURAL NETWORK TRAINING		
Attorney Docket Number	P-587424-US	Small Entity Status Claimed ☒	
Application Type	Nonprovisional		
Subject Matter	Utility		
Total Number of Drawing Sheets (if any)	8	Suggest Figure for Publication (if any)	

Filing By Reference:

Only complete this section when filing an application by reference under 35 U.S.C. 111(c) and 37 CFR 1.57(a). Do not complete this section if application papers including a specification and any drawings are being filed. Any domestic benefit or foreign priority information must be provided in the appropriate section(s) below (i.e., "Domestic Benefit/National Stage Information" and "Foreign Priority Information").

For the purposes of a filing date under 37 CFR 1.53(b), the description and any drawings of the present application are replaced by this reference to the previously filed application, subject to conditions and requirements of 37 CFR 1.57(a).

Application number of the previously filed application	Filing date (YYYY-MM-DD)	Intellectual Property Authority or Country

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76		Attorney Docket Number	P-587424-US
		Application Number	Not yet assigned
Title of Invention	SYSTEMS AND METHODS FOR PREDICTIVE EARLY STOPPING IN NEURAL NETWORK TRAINING		

Publication Information:

☐	Request Early Publication (Fee required at time of Request 37 CFR 1.219)
☐	Request Not to Publish. I hereby request that the attached application not be published under 35 U.S.C. 122(b) and certify that the invention disclosed in the attached application has not and will not be the subject of an application filed in another country, or under a multilateral international agreement, that requires publication at eighteen months after filing.

Representative Information:

Representative information should be provided for all practitioners having a power of attorney in the application. Providing this information in the Application Data Sheet does not constitute a power of attorney in the application (see 37 CFR 1.32). Either enter Customer Number or complete the Representative Name section below. If both sections are completed the customer Number will be used for the Representative Information during processing.			
Please Select One: ☒ Customer Number ☐ US Patent Practitioner ☐ Limited Recognition (37 CFR 11.9)			
Customer Number	49443		

Domestic Benefit/National Stage Information:

This section allows for the applicant to either claim benefit under 35 U.S.C. 119(e), 120, 121, or 365(c) or indicate National Stage entry from a PCT application. Providing this information in the application data sheet constitutes the specific reference required by 35 U.S.C. 119(e) or 120, and 37 CFR 1.78.			
When referring to the current application, please leave the "Application Number" field blank.			
Prior Application Status	Pending		
Application Number	Continuity Type	Prior Application Number	Filing or 371(c) Date (YYYY-MM-DD)
	Claims benefit of provisional	62/852,525	2019-05-24
Prior Application Status			
Application Number	Continuity Type	Prior Application Number	Filing or 371(c) Date (YYYY-MM-DD)
Additional Domestic Benefit/National Stage Data may be generated within this form by selecting the Add button. 			

Foreign Priority Information:

This section allows for the applicant to claim priority to a foreign application. Providing this information in the application data sheet constitutes the claim for priority as required by 35 U.S.C. 119(b) and 37 CFR 1.55(d). When priority is claimed to a foreign application that is eligible for retrieval under the priority document exchange program (PDX) the information will be used by the Office to automatically attempt retrieval pursuant to 37 CFR 1.55(h)(1) and (2). Under the PDX program, applicant bears the ultimate responsibility for ensuring that a copy of the foreign application is received by the Office from the participating foreign intellectual property office, or a certified copy of the foreign priority application is filed, within the time period specified in 37 CFR 1.55(g)(1).			
			
Application Number	Country	Filing Date (YYYY-MM-DD)	Access Code (if applicable)
Additional Foreign Priority Data may be generated within this form by selecting the Add button. 			

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76		Attorney Docket Number	P-587424-US
		Application Number	Not yet assigned
Title of Invention	SYSTEMS AND METHODS FOR PREDICTIVE EARLY STOPPING IN NEURAL NETWORK TRAINING		

Statement under 37 CFR 1.55 or 1.78 for AIA (First Inventor to File) Transition Applications

☐	This application (1) claims priority to or the benefit of an application filed before March 16, 2013 and (2) also contains, or contained at any time, a claim to a claimed invention that has an effective filing date on or after March 16, 2013. NOTE: By providing this statement under 37 CFR 1.55 or 1.78, this application, with a filing date on or after March 16, 2013, will be examined under the first inventor to file provisions of the AIA.
--------------------------	---

Authorization or Opt-Out of Authorization to Permit Access:

When this Application Data Sheet is properly signed and filed with the application, applicant has provided written authority to permit a participating foreign intellectual property (IP) office access to the instant application-as-filed (see paragraph A in subsection 1 below) and the European Patent Office (EPO) access to any search results from the instant application (see paragraph B in subsection 1 below). Should applicant choose not to provide an authorization identified in subsection 1 below, applicant must opt-out of the authorization by checking the corresponding box A or B or both in subsection 2 below. NOTE: This section of the Application Data Sheet is ONLY reviewed and processed with the INITIAL filing of an application. After the initial filing of an application, an Application Data Sheet cannot be used to provide or rescind authorization for access by a foreign IP office(s). Instead, Form PTO/SB/39 or PTO/SB/69 must be used as appropriate.
1. Authorization to Permit Access by a Foreign Intellectual Property Office(s) A. Priority Document Exchange (PDX) - Unless box A in subsection 2 (opt-out of authorization) is checked, the undersigned hereby <u>grants the USPTO authority</u> to provide the European Patent Office (EPO), the Japan Patent Office (JPO), the Korean Intellectual Property Office (KIPO), the State Intellectual Property Office of the People's Republic of China (SIPO), the World Intellectual Property Organization (WIPO), and any other foreign intellectual property office participating with the USPTO in a bilateral or multilateral priority document exchange agreement in which a foreign application claiming priority to the instant patent application is filed, access to: (1) the instant patent application-as-filed and its related bibliographic data, (2) any foreign or domestic application to which priority or benefit is claimed by the instant application and its related bibliographic data, and (3) the date of filing of this Authorization. See 37 CFR 1.14(h)(1). B. Search Results from U.S. Application to EPO - Unless box B in subsection 2 (opt-out of authorization) is checked, the undersigned hereby <u>grants the USPTO authority</u> to provide the EPO access to the bibliographic data and search results from the instant patent application when a European patent application claiming priority to the instant patent application is filed. See 37 CFR 1.14(h)(2). The applicant is reminded that the EPO's Rule 141(1) EPC (European Patent Convention) requires applicants to submit a copy of search results from the instant application without delay in a European patent application that claims priority to the instant application.
2. Opt-Out of Authorizations to Permit Access by a Foreign Intellectual Property Office(s) ☐ A. Applicant DOES NOT authorize the USPTO to permit a participating foreign IP office access to the instant application-as-filed. If this box is checked, the USPTO will not be providing a participating foreign IP office with any documents and information identified in subsection 1A above. ☐ B. Applicant DOES NOT authorize the USPTO to transmit to the EPO any search results from the instant patent application. If this box is checked, the USPTO will not be providing the EPO with search results from the instant application. NOTE: Once the application has published or is otherwise publicly available, the USPTO may provide access to the application in accordance with 37 CFR 1.14.

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	P-587424-US
	Application Number	Not yet assigned
Title of Invention	SYSTEMS AND METHODS FOR PREDICTIVE EARLY STOPPING IN NEURAL NETWORK TRAINING	

Applicant Information:

Providing assignment information in this section does not substitute for compliance with any requirement of part 3 of Title 37 of CFR to have an assignment recorded by the Office.							
Applicant 1							
If the applicant is the inventor (or the remaining joint inventor or inventors under 37 CFR 1.45), this section should not be completed. The information to be provided in this section is the name and address of the legal representative who is the applicant under 37 CFR 1.43; or the name and address of the assignee, person to whom the inventor is under an obligation to assign the invention, or person who otherwise shows sufficient proprietary interest in the matter who is the applicant under 37 CFR 1.46. If the applicant is an applicant under 37 CFR 1.46 (assignee, person to whom the inventor is obligated to assign, or person who otherwise shows sufficient proprietary interest) together with one or more joint inventors, then the joint inventor or inventors who are also the applicant should be identified in this section.							
							
☒ Assignee	☒ Legal Representative under 35 U.S.C. 117	☒ Joint Inventor					
☒ Person to whom the inventor is obligated to assign.		☒ Person who shows sufficient proprietary interest					
If applicant is the legal representative, indicate the authority to file the patent application, the inventor is:							
Name of the Deceased or Legally Incapacitated Inventor :							
If the Assignee is an Organization check here. ☒							
Organization Name	Comet ML, Inc.						
Prefix	Given Name	Middle Name	Family Name	Suffix			
Mailing Address Information For Applicant:							
Address 1	100 Avenue of the Americas						
Address 2	3rd Floor						
City	New York	State/Province	NY				
Country	USA	Postal Code	10013				
Phone Number		Fax Number					
Email Address							
Additional Applicant Data may be generated within this form by selecting the Add button. 							

Assignee Information including Non-Applicant Assignee Information:

Providing assignment information in this section does not substitute for compliance with any requirement of part 3 of Title 37 of CFR to have an assignment recorded by the Office.				
Assignee 1				
Complete this section if assignee information, including non-applicant assignee information, is desired to be included on the patent application publication. An assignee-applicant identified in the "Applicant Information" section will appear on the patent application publication as an applicant. For an assignee-applicant, complete this section only if identification as an assignee is also desired on the patent application publication.				
Prefix	Given Name	Middle Name	Family Name	Suffix

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	P-587424-US
	Application Number	Not yet assigned
Title of Invention	SYSTEMS AND METHODS FOR PREDICTIVE EARLY STOPPING IN NEURAL NETWORK TRAINING	

If the Assignee or Non-Applicant Assignee is an Organization check here. ☒			
Organization Name		Comet ML, Inc.	
Mailing Address Information For Assignee including Non-Applicant Assignee:			
Address 1		100 Avenue of the Americas	
Address 2		3rd Floor	
City		New York	State/Province NY
Country	USA	Postal Code	10013
Phone Number		Fax Number	
Email Address			
Additional Assignee or Non-Applicant Assignee Data may be generated within this form by selecting the Add button. Add			

Signature:

NOTE: This Application Data Sheet must be signed in accordance with 37 CFR 1.33(b). **However, if this Application Data Sheet is submitted with the INITIAL filing of the application and either box A or B is not checked in subsection 2 of the "Authorization or Opt-Out of Authorization to Permit Access" section, then this form must also be signed in accordance with 37 CFR 1.14(c).**

This Application Data Sheet **must** be signed by a patent practitioner if one or more of the applicants is a **juristic entity** (e.g., corporation or association). If the applicant is two or more joint inventors, this form must be signed by a patent practitioner, all joint inventors who are the applicant, or one or more joint inventor-applicants who have been given power of attorney (e.g., see USPTO Form PTO/AIA/81) on behalf of **all** joint inventor-applicants.

See 37 CFR 1.4(d) for the manner of making signatures and certifications.

Signature	/Caleb Pollack/		Date (YYYY-MM-DD)	2019-12-09
First Name	Caleb	Last Name	Pollack	Registration Number 37,912
Additional Signature may be generated within this form by selecting the Add button. Add				

This collection of information is required by 37 CFR 1.76. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 23 minutes to complete, including gathering, preparing, and submitting the completed application data sheet form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Substitute for form 1449B/PTO INFORMATION DISCLOSURE STATEMENT BY APPLICANT <i>(use as many sheets as necessary)</i> <i>(use as many sheets as necessary)</i>				Complete if Known	
				Application Number	Not yet assigned
				Filing Date	Herewith
				First Named Inventor	Dhruv NAIR
				Art Unit	Not yet assigned
				Examiner Name	Not yet assigned
Sheet	1	of	1	Attorney Docket Number	P-587424-US

NON PATENT LITERATURE DOCUMENTS			
Examiner Initials*	Cite No. ¹	Include name of the author (in CAPITAL LETTERS), title of the article (where appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published.	T ²
		BAKER et al., "Accelerating Neural Architecture Search Using Performance Prediction", arXiv:1705.10823v2 [cs.LG] November 8, 2017.	☐
		SWERSKY et al., "Freeze-Thaw Bayesian Optimization", arXiv:1406.3896v1 [stat.ML] June 16, 2014.	☐
		KLEIN et al., "Learning Curve Prediction with Bayesian Neural Networks", 2016.	☐
		DOMHAN et al., "Speeding up Automatic Hyperparameter Optimization of Deep Neural Networks by Extrapolation of Learning Curves", Twenty-Fourth International Joint Conference on Artificial Intelligence, 2015.	☐
			☐
			☐
			☐
			☐
			☐
			☐
			☐

Examiner Signature		Date Considered	
-----------------------	--	--------------------	--

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant.
¹ Applicant's unique citation designation number (optional). ² Applicant is to place a check mark here if English language Translation is attached.
 This collection of information is required by 37 CFR 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 2 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 (1-800-786-9199) and select option 2.

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Inventors:	Dhruv NAIR et al.	Examiner:	Not yet assigned
Serial No.:	Not yet assigned	Group Art Unit:	Not yet assigned
Filed:	Herewith	Confirmation No.:	Not yet assigned
Title:	SYSTEMS AND METHODS FOR PREDICTIVE EARLY STOPPING IN NEURAL NETWORK TRAINING		

Commissioner for Patents
P. O. Box 1450
Alexandria, VA 22313-1450

Sir/Madam:

INFORMATION DISCLOSURE STATEMENT

Pursuant to 37 C.F.R. §§1.56, 1.97 and 1.98, this Information Disclosure Statement includes Form PTO/SB/08:

1. ☐ listing documents including patents, publications and other information for consideration by the Examiner, however, since the subject application was filed after June 30, 2003, copies of United States patents and/or United States patent application publications are not included in this information disclosure statement; and/or
2. ☐ listing documents including patents, publications and other information that have been previously cited or submitted to the Patent Office in prior application U.S. Patent application No. _____, filed _____, which is properly identified and relied on for an earlier effective filing date under 35 U.S.C. 120 for consideration by the Examiner; however, in accordance with 37 C.F.R. 1.98(d), copies of such documents are not included in this information disclosure statement; and/or
3. ☒ listing documents including patents, publications, and other information for consideration by the Examiner, copies of which are included with this information disclosure statement;
4. ☐ listing documents including patents, publications, and other information for consideration by the Examiner, which were cited in a communication from a foreign patent office in a counterpart foreign application, copies of which are included with this information disclosure statement.

The information herein cited is only in fulfillment of Applicant(s) duty of candor in disclosing all information brought to Applicant(s) attention. This submission does not represent that a search has been made or that no better art exists and does not constitute an admission that each or all of the listed documents are material or constitute "prior art". If it should be determined that any of the listed documents do not constitute

APPLICANTS: Dhruv NAIR et al.
SERIAL NO.: Not yet assigned
FILED: Herewith
Page 2

"prior art" under United States law, Applicant(s) reserve the right to present to the office the relevant facts and law regarding the appropriate status of such documents.

Applicant(s) further reserve(s) the right to take appropriate action to establish the patentability of the disclosed invention over the listed documents, should one or more of the documents be applied against the claims of the present application.

In accordance with MPEP Sections 609 and 707.05(b), it is requested that each and every document cited (including any cited in applicant's specification which is not repeated on the attached Form PTO-1449) be given thorough consideration and that it be cited of record in the prosecution history of the present application by initialing on Form PTO-1449. Such initialing is requested even if the Examiner does not consider it to be prior art for any reason, or even if the Examiner does not believe that the guidelines for citation have been fully complied with. This is requested so that each document becomes listed on the face of the patent issuing on the present application and is evidence that the Examiner has considered the document. Neither the submission of the Information Disclosure Statement nor the content of any document cited herein constitutes a disclaimer of claim scope or an admission that any item submitted constitutes prior art.

This Information Disclosure Statement is being filed:

- I) ☒ Within three (3) months of filing the subject Application or entry of the subject Application into the national stage; or before mailing of the first Office Action on the merits of the subject Application, whichever event occurs last; or before mailing of the first Office Action on the merits after filing a request for continued examination, pursuant to of 37 C.F.R §1.97 (b); or
- II) ☐ After the period specified in (I) but before the mailing date of either a final Official Action under 37 C.F.R §1.113 or a notice of allowance under 37 C.F.R §1.311 whichever occurs first and;

1. ☐ the undersigned hereby states that each item of information listed on the Form PTO-1449 was either (i) cited in a communication from a foreign patent office in a counterpart foreign application not more than three (3) months prior to the filing of this Information Disclosure Statement or (ii) not cited in a communication from a foreign patent office in a counterpart foreign application, and, to the knowledge of the undersigned after making reasonable inquiry, not known to any individual designated in §1.56(c) more than three (3) months prior to the filing of this information disclosure statement; or

2. ☐ the fee in the amount of \$240 (Large Entity) under 37 C.F.R §1.17 (p) is being paid by credit card (note deposit account information below in the event of lack of credit card payment).

- III) ☐ After the period in (I) and (II) but before the payment of the issue fee and,
1. The undersigned hereby states:

APPLICANTS: Dhruv NAIR et al.
SERIAL NO.: Not yet assigned
FILED: Herewith
Page 3

- a) ☐ that each item of information cited on the form PTO-1449 was first cited in a communication from a foreign Patent Office in a counterpart foreign application not more than three (3) months prior to the filing of this Information Disclosure Statement; or
- b) ☐ that no items of information contained in Form PTO-1449 were cited in a communication from a foreign patent office in a counterpart foreign application, and to the knowledge of the undersigned after making reasonable inquiry, no item of information contained in this Information Disclosure Statement was known to any individual designated in 37 C.F.R. § 1.56(c) more than three months prior to the filing of this Information Disclosure Statement; and
2. ☐ The Petition fee in the amount of \$240 (Large Entity) under 37 C.F.R §1.17 is being paid by credit card (note deposit account information below in the event of lack of credit card payment).

To the extent that there are any fees associated with this filing that are due but are not paid or are not able to be paid by a credit card mentioned elsewhere in this document, Applicants authorize the U.S. Patent and Trademark Office to charge any such fees to deposit account No. **50-3355**.

Respectfully submitted,

/Caleb Pollack/
Caleb Pollack
Attorney/Agent for Applicant(s)
Registration No. 37,912

Dated: December 9, 2019

Pearl Cohen Zedek Latzer Baratz LLP
1500 Broadway, 12th Floor
New York, New York 10036
Tel: (646) 878-0800
Fax: (646) 878-0801

UTILITY PATENT APPLICATION TRANSMITTAL <i>(Only for new nonprovisional applications under 37 CFR 1.53(b))</i>	Attorney Docket No.	P-587424-US
	First Inventor	Dhruv NAIR
	Title	SYSTEMS AND METHODS FOR PREDICTIVE EARLY STOPPING IN NEURAL NETWORK TRAINING
	Express Mail Label No.	

APPLICATION ELEMENTS <i>See MPEP chapter 600 concerning patent application contents.</i>	ADDRESS TO: Commissioner for Patents P. O. Box 1450 Alexandria, VA 22313-1450
1. ☐ Fee Transmittal Form (PTO/SB/17 or equivalent) 2. ☒ Applicant asserts small entity status. See 37 CFR 1.27 3. ☐ Applicant certifies micro entity status. See 37 CFR 1.29. Applicant must attach form PCT/SB/15A or B or equivalent. 4. ☒ Specification [Total Pages 34] Both the claims and abstract must start on a new page. (See MPEP § 608.01(a) for information on the preferred arrangement) 5. ☒ Drawing(s) (35 U.S.C. 113) [Total Sheets 8] 6. Inventor's Oath or Declaration [Total Pages 3] (including substitute statements under 37 CFR 1.64 and assignments serving as an oath or declaration under 37 CFR 1.63(e)) a. ☒ Newly Executed (original or copy) b. ☐ A copy from a prior application (37 C.F.R. 1.63(d)) 7. ☒ Application Data Sheet. *See note below. See 37 CFR 1.76 (PTO/AIA/14 or equivalent) 8. CD-ROM or CD-R in duplicate, large table or Computer Program (Appendix) ☐ Landscape Table on CD 9. Nucleotide and/or Amino Acid Sequence Submission (if applicable, items a. - c. are required) a. ☐ Computer Readable Form (CRF) b. ☐ Specification Sequence Listing on: i. ☐ CD-ROM or CD-R (2 copies); or ii. ☐ Paper c. ☐ Statements verifying identity of above copies	ACCOMPANYING APPLICATION PAPERS 10. ☐ Assignment Papers (cover sheet & document(s)) Name of Assignee _____ 11. ☒ 37 CFR 3.73(c) Statement ☒ Power of Attorney (when there is an assignee) 12. ☐ English Translation Document (if applicable) 13. ☒ Information Disclosure Statement (PTO/SB/08 or PTO-1449) ☒ Copies of citations attached 14. ☐ Preliminary Amendment 15. ☐ Return Receipt Postcard (MPEP § 503) (Should be specifically itemized) 16. ☐ Certified Copy of Priority Document(s) (if foreign priority is claimed) 17. ☐ Nonpublication Request Under 35 U.S.C. 122(b)(2)(B)(i). Applicant must attach form PTO/SB/35 or equivalent. 18. ☒ Other: <u>1) Petition Under 37 CFR 1.84(a)(2) to Accept Color Drawings</u> <u>2) Authorization for Internet Communications in a Patent Application</u>

*Note: (1) Benefit claims under 37 CFR 1.78 and foreign priority claims under 1.55 **must** be included in an Application Data Sheet (ADS).
(2) For applications filed under 35 U.S.C. 111, the application must contain an ADS specifying the applicant if the applicant is an assignee, person to whom the inventor is under an obligation to assign, or person who otherwise shows sufficient proprietary interest in the matter. See 37 CFR 1.46(b).

19. PAYMENT			
To the extent that there are any fees associated with this filing that are due but are not paid or are not able to be paid by a credit card mentioned elsewhere in this document, Applicants authorize the U.S. Patent and Trademark Office to charge any such fees to Deposit Account No. 50-3355			
20. CORRESPONDENCE ADDRESS			
☒ The address associated with Customer Number: 49443		or ☐ Correspondence address below	
Name	Pearl Cohen Zedek Latzer Baratz LLP		
Address	1500 Broadway, 12th Floor		
City	New York	State	New York
Country	USA	Zip Code	10036
	Telephone	(646) 878-0800	Email
			USPTO@PearlCohen.com

Signature	/Caleb Pollack/	Date	9 December 2019
Name (Print/Type)	Caleb Pollack	Registration No. (Attorney/Agent)	37,912

This collection of information is required by 37 CFR 1.53(b). The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.11 and 1.14. This collection is estimated to take 12 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

Electronic Patent Application Fee Transmittal				
Application Number:				
Filing Date:				
Title of Invention:		SYSTEMS AND METHODS FOR PREDICTIVE EARLY STOPPING IN NEURAL NETWORK TRAINING		
First Named Inventor/Applicant Name:		Dhruv NAIR		
Filer:		Caleb Pollack/Alex Matt		
Attorney Docket Number:		P-587424-US		
Filed as Small Entity				
Filing Fees for Utility under 35 USC 111(a)				
Description	Fee Code	Quantity	Amount	Sub-Total in USD(\$)
Basic Filing:				
UTILITY FILING FEE (ELECTRONIC FILING)	4011	1	75	75
UTILITY SEARCH FEE	2111	1	330	330
UTILITY EXAMINATION FEE	2311	1	380	380
Pages:				
Claims:				
CLAIMS IN EXCESS OF 20	2202	2	50	100
INDEPENDENT CLAIMS IN EXCESS OF 3	2201	1	230	230
Miscellaneous-Filing:				

Description	Fee Code	Quantity	Amount	Sub-Total in USD(\$)
Petition:				
PETITION FEE-37CFR 1.17(H) (GROUP II)	2464	1	70	70
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				
Extension-of-Time:				
Miscellaneous:				
Total in USD (\$)				1185

Electronic Acknowledgement Receipt	
EFS ID:	37966269
Application Number:	16707265
International Application Number:	
Confirmation Number:	1015
Title of Invention:	SYSTEMS AND METHODS FOR PREDICTIVE EARLY STOPPING IN NEURAL NETWORK TRAINING
First Named Inventor/Applicant Name:	Dhruv NAIR
Customer Number:	49443
Filer:	Caleb Pollack/Alex Matt
Filer Authorized By:	Caleb Pollack
Attorney Docket Number:	P-587424-US
Receipt Date:	09-DEC-2019
Filing Date:	
Time Stamp:	14:05:28
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	yes
Payment Type	CARD
Payment was successfully received in RAM	\$1185
RAM confirmation Number	E2019B9E05501054
Deposit Account	503355
Authorized User	Alex Matt
The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows: 37 CFR 1.16 (National application filing, search, and examination fees) 37 CFR 1.17 (Patent application and reexamination processing fees)	

37 CFR 1.19 (Document supply fees)					
37 CFR 1.20 (Post Issuance fees)					
37 CFR 1.21 (Miscellaneous fees and charges)					
File Listing:					
Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1		P-587424-US-U-APP-09DEC19.pdf	2027952	yes	70
			a1eaa7990958f89950c8e4934e898056c9e407e		
	Multipart Description/PDF files in .zip description				
	Document Description		Start	End	
	Information Disclosure Statement (IDS) Form (SB08)		70	70	
	Transmittal Letter		67	69	
	Internet Communications Authorized		66	66	
	Miscellaneous Incoming Letter		64	65	
	Assignee showing of ownership per 37 CFR 3.73		59	63	
	Power of Attorney		58	58	
	Oath or Declaration filed		55	57	
	Drawings-other than black and white line drawings		47	54	
	Abstract		46	46	
	Claims		42	45	
	Specification		8	41	
	Application Data Sheet		2	7	
	Transmittal of New Application		1	1	
Warnings:					
Information:					

2	Non Patent Literature	BAKERetal-2017.pdf	696064	no	14
			2cad68ca9cea089324b67ed96b1175f7a655c521		
Warnings:					
Information:					
3	Non Patent Literature	SWERSKYetal-2014.pdf	1010029	no	12
			13180ee816b8aaae09072cc97de4e12d86732841		
Warnings:					
Information:					
4	Non Patent Literature	KLEINetal-2016.pdf	6808831	no	16
			42fa2c1cf0213d491e9af192b08b188132de6a99		
Warnings:					
Information:					
5	Non Patent Literature	DOMHANetal-2015.pdf	2449627	no	9
			efb4431579a46d9cfa51b1e1bbd4ddb9f4a30246		
Warnings:					
Information:					
6	Fee Worksheet (SB06)	fee-info.pdf	39964	no	2
			b9b02a1120e90ac10e875d060a42007e1bd b894c		
Warnings:					
Information:					
Total Files Size (in bytes):			13032467		

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.