

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

GOOGLE LLC
Petitioner

v.

K.MIZRA LLC
Patent Owner

Case No. IPR2026-00304
Patent 8,438,120

**DECLARATION OF MICHAEL J. PAZZANI IN SUPPORT OF
PETITION FOR *INTER PARTES* REVIEW
OF U.S. PATENT NO. 8,438,120**

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

TABLE OF CONTENTS

I.	INTRODUCTION	1
II.	BACKGROUND AND QUALIFICATIONS	1
III.	INFORMATION CONSIDERED	6
IV.	RELEVANT LEGAL STANDARDS	6
A.	Claim Interpretation	6
B.	Perspective of One of Ordinary Skill in the Art.....	7
C.	Anticipation	8
D.	Obviousness.....	8
V.	SUMMARY OF OPINIONS.....	11
VI.	TECHNOLOGY BACKGROUND.....	12
VII.	THE CHALLENGED '120 PATENT.....	16
VIII.	THE '120 PATENT PROSECUTION HISTORY	17
IX.	LEVEL OF ORDINARY SKILL IN THE ART	17
X.	PRIORITY DATE	18
XI.	CLAIM CONSTRUCTION	19
XII.	OVERVIEW OF THE PRIOR ART REFERENCES	20
A.	Watanabe (EX1006).....	20
B.	Friedrichs (EX1011).....	26
C.	de Boer (EX1007)	27
D.	Kirsten (EX1008)	28

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

E.	Xu (EX1017)	29
XIII.	DETAILED EXPLANATION OF THE UNPATENTABILITY GROUNDS	29
A.	Claims 1-8 and 11-16 Are Anticipated by Watanabe (EX1006), And/Or Would Have Been Obvious Over Watanabe (EX1006), And/Or Would Have Been Obvious Over Watanabe in View of Friedrichs (Ground 1).....	30
1.	Claim 1 (Preamble): “A method of determining hyperparameters of a classifier in a machine learning system by iteratively producing an estimate of a target hyperparameter vector, each iteration comprising the steps of:”	30
a.	[1.1]: “drawing a random sample of hyperparameter vectors from a set of possible hyperparameter vectors,”	46
b.	[1.2]: “updating the estimate of the target hyperparameter vector by using the random sample, and”	47
c.	[1.3]: “selecting, from the random sample of hyperparameter vectors, a hyperparameter vector producing a best result in the present and any previous iterations, and”	52
d.	[1.4]: “wherein the step of updating the estimate of the target hyperparameter vector uses said hyperparameter vector producing the best result.”	55
2.	Claim 2: “The method according to claim 1, further comprising the steps of: selecting a further hyperparameter vector producing the best result in any previous iterations, and restricting the random sample of hyperparameter vectors by using the further selected hyperparameter vector.”	56

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

3. Claim 3: “The method according to claim 2, wherein the step of restricting the random sample of hyperparameter vectors involves using an interval surrounding the further selected hyperparameter vector.” 57
4. Claim 4: “The method according to claim 3, wherein the step of restricting the random sample of hyperparameter vectors is carried out prior to the step of selecting the hyperparameter vector producing the best result in the present and any previous iterations.” 58
5. Claim 5: “The method according to claim 3, wherein the step of restricting the random sample of hyperparameter vectors is carried out during the step of updating the estimate of the target hyperparameter vector.” 59
6. Claim 6: “The method according to claim 2, wherein the step of restricting the random sample of hyperparameter vectors is carried out prior to the step of selecting the hyperparameter vector producing the best result in the present and any previous iterations.” 60
7. Claim 7: “The method according to claim 2, wherein the step of restricting the random sample of hyperparameter vectors is carried out during the step of updating the estimate of the target hyperparameter vector.” 61
8. Claim 8: “The method according to claim 1, wherein the step of updating the estimate of the hyperparameter vector uses a weighting function.” 62
9. Claim 11: “The method according to claim 1, wherein the steps are carried out by a programmed computer apparatus including a processor and a computer readable medium including computer executable instructions.” 63
10. Claim 12: “A classifier for use in a machine learning system, the classifier using hyperparameters as control parameters, wherein the hyperparameters are determined according to the set of steps recited in claim 1.” 64

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

11. Claim 13 (Preamble): “A non-transitory computer readable medium product including computer-executable instructions for carrying out a method of determining hyperparameters of a classifier in a machine learning system by iteratively producing an estimate of a target hyperparameter vector, each iteration comprising the steps of:” 65
 - a. [13.1]: “drawing a random sample of hyperparameter vectors from a set of possible hyperparameter vectors,” 65
 - b. [13.2]: “updating the estimate of the target hyperparameter vector by using the random sample, and” 65
 - c. [13.3]: “selecting, from the random sample of hyperparameter vectors, a hyperparameter vector producing a best result in the present and any previous iterations, and” 66
 - d. [13.4]: “wherein the step of updating the estimate of the target hyperparameter vector uses said hyperparameter vector producing the best result.” 66
12. Claim 14 (Preamble): “A device for determining hyperparameters of a classifier in a machine learning system, the device comprising a processor arranged for iteratively producing an estimate of a target hyperparameter vector, each iteration comprising the steps of:” 66
 - a. [14.1]: “drawing a random sample of hyperparameter vectors from a set of possible hyperparameter vectors,” 66
 - b. [14.2]: “updating the estimate of the hyperparameter vector by using the random sample, and” 67

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

- c. [14.3]: “selecting, from the random sample of hyperparameter vectors, a hyperparameter vector producing a best result in the present and any previous iterations, and” 67
 - d. [14.4]: “wherein the step of updating the estimate of the target hyperparameter vector uses said hyperparameter vector producing the best result.” 67
 - 13. Claim 15: “The device according to claim 14, wherein the processor is further arranged for: selecting a further hyperparameter vector producing the best result in any previous iterations, and restricting the random sample by using the hyperparameter vector producing the best result in any previous iteration.” 67
 - 14. Claim 16: “A machine learning system, comprising the device for determining hyperparameters of a classifier defined in claim 14.” 67
- B. Claim 9 Would Have Been Obvious Over Watanabe (EX1006), in View of de Boer (EX1007) (Ground 2) 68
 - 1. Claim 9: “The method according to claim 1, wherein E^t is the selected hyperparameter vector X^t_i at iteration t producing the best result $S(X^t_i)$, the selected hyperparameter vector X^t_i having elements X^t_{ij} , wherein the step of updating the estimate of the target hyperparameter vector comprises the step of determining

$$v^t_j = \frac{\sum_{i=1}^n I\{S(X^t_i) \geq \gamma^t\} W(X^t_i; E^t) X^t_{ij}}{\sum_{i=1}^n I\{S(X^t_i) \geq \gamma^t\} W(X^t_i; E^t)},$$
 the hyperparameter v^t_j , where wherein γ^t is a threshold value and W is a weighting function.” 68
- C. Claim 10 Would Have Been Obvious Over Watanabe (EX1006), in View of de Boer (EX1007), Kirsten (EX1008), and Xu (EX1017) (Ground 3) 74

1. Claim 10: “The method according to claim 9, wherein the weighting function W is given by

XIV. SECONDARY CONSIDERATIONS.....80

XV. CONCLUSION.....81

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

LIST OF EXHIBITS

Exhibit No.	Description
1001	U.S. Patent No. 8,438,120, issued May 7, 2013
1002	File History of U.S. Patent Application No. 12/597,257
1003	PCT Publication No. WO 2008/133509, published on November 6, 2008
1004	EP 07112037.2, filed on July 9, 2007
1005	EP 07106963.7, filed on April 25, 2007
1006	U.S. Patent No. 6,004,015, issued on December 21, 1999 to T. Watanabe <i>et al.</i> (“Watanabe”)
1007	Pieter-Tjerk De de Boer <i>et al.</i> , “A tutorial on the Cross-Entropy Method,” ANNALS OF OPERATIONS RESEARCH 134:19-67 (2005) (“de Boer”)
1008	M. Kirsten <i>et al.</i> , “Distance Based Approaches to Relational Learning and Clustering,” in Relational Data Mining, Chpt. 9, pp. 213-232 (2001) (“Kirsten”)
1009	File History of EP 08741668.1
1010	S. Mannor, “The Cross Entropy Method For Classification,” pp. 561-568 (2005) (“Mannor”)
1011	F. Friedrichs, <i>et al.</i> , “Evolutionary tuning of multiple SVM parameters,” pp107-117 (January 7, 2005) (“Friedrichs”)
1012	U.S. Patent No. 5,140,530, issued on August 18, 1992 to A. Guha <i>et al.</i> (“Guha”)
1013	U.S. Patent No. 5,255,345, issued October 19, 1993 to Craig G. Shaefer (“Shaefer”)

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

Exhibit No.	Description
1014	T. Horváth <i>et al.</i> , “Relational Instance-Based Learning with Lists and Terms,” MACHINE LEARNING 43:53-80 (2001) (“Horvath”)
1015	S. Raaijmakers, “Sentiment Classification with Interpolated Information Diffusion Kernels.” TNO Information and Communication Technology, 36-39 (2007) (“Raaijmakeers”).
1016	R. Caruana, <i>et al.</i> (2000). "Overfitting in neural nets: Backpropagation, conjugate gradient, and early stopping." Advances in neural information processing systems 13 (“Caruana”).
1017	R. Xu, <i>et al.</i> "Survey of clustering algorithms." IEEE Transactions on neural networks 16, no. 3 (2005): 645-678. (“Xu”).
1019	<i>Curriculum Vitae</i> of Michael J. Pazzani, Ph.D.
1022	Complaint filed in <i>Google LLC v. K.Mizra LLC</i> , Civ. Action No. 3:25-cv-08107 (N.D. Cal.) on September 24, 2025
1023	Plaintiff’s Notice of Voluntary Dismissal Without Prejudice filed in Civ. Action No. 3:25-cv-08107 (N.D. Cal.) on March 27, 2026
1024	Complaint filed in <i>K.Mizra LLC v. Google LLC</i> , Civ. Action No. 1:26-cv-00754 (W.D. Tex.) on March 27, 2026
1025	U.S. Patent No. 5,249,259, issued on September 28, 1993 to R. Harvey (“Harvey”)

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

I, Michael J. Pazzani, hereby state as follows:

I. INTRODUCTION

1. I have been retained by Google LLC (“Petitioner”) as an independent expert consultant in this *inter partes* review (“IPR”) proceeding before the United States Patent and Trademark Office (“PTO”).

2. I have been asked by Counsel for Petitioners (“Counsel”) to consider whether certain references teach or suggest the features recited in Claims 1-16 (“the Challenged Claims”) of U.S. Patent No. 8,438,120 (“the ’120 Patent”) (EX1001). My opinions and the bases for my opinions are set forth below.

3. I am being compensated at my ordinary and customary consulting rate for my work, which is \$650 per hour. My compensation is in no way contingent on the nature of my findings, the presentation of my findings in testimony, or the outcome of this or any other proceeding. I have no other financial interest in this proceeding.

II. BACKGROUND AND QUALIFICATIONS

4. All of my opinions stated in this declaration are based on my own personal knowledge and professional judgment. In forming my opinions, I have relied on my knowledge and experience in designing, developing, researching, and teaching the technology referenced in this declaration.

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

5. I am over 18 years of age and, if I am called upon to do so, I would be competent to testify as to the matters set forth herein. I understand that a copy of my current curriculum vitae, which details my education and professional and academic experience, is being submitted as EX1019. The following provides a brief overview of some of my experience that is relevant to the matters set forth in this declaration.

6. I have over 45 years of experience in the field of information and computer science.

7. I am presently an Affiliated Scientist at the Information Sciences Institute at the University of Southern California, having retired on October 4, 2024, from a position as a Distinguished Principal Scientist. I have worked in this role since 2021. From 2019 to 2022, I was a Distinguished Scientist at the Halicioğlu Data Science Institute at the University of California, San Diego. From 2012 to 2019, I was a professor of Computer Science and Engineering and the Vice Chancellor for Research and Economic Development at the University of California, Riverside. From 2006 to 2012, I was a distinguished professor of Computer Science and the Vice President for Research and Economic Development at Rutgers University. From 2002 to 2006, I worked as the Division Director of Information and Intelligent Systems at the National Science Foundation (“NSF”), a federal agency that funds fundamental research at universities. This division, with an annual

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

budget of approximately \$250 million, provides funding for research on topics that included Internet Search, Natural Language Processing, Artificial Intelligence, Digital Libraries, Machine Learning, Robotics, Computer Vision, Human-Computer Interaction, and Neural Networks.

8. From 1988 to 2005, I was a professor of Information and Computer Science at the University of California, Irvine (“UC Irvine”), where I taught courses in Artificial Intelligence, Information Retrieval, Search Engines, and Machine Learning. I conducted research and supervised graduate students on projects that included Artificial Intelligence, Internet Search, Personalization and Recommendation Systems, Neural Networks, and Machine Learning. In 1999, while on leave from UC Irvine, I was CEO and co-founder of AdaptiveInfo.com (“AdaptiveInfo”), a company that personalized search and delivery of Internet news, classified ads, and restaurant listings for mobile devices such as cell phones. AdaptiveInfo was sold to Versifi Technologies in 2003.

9. I am an author or co-author of 52 journal articles, 85 conference papers, 14 book chapters and one book. I have also presented at approximately 32 workshops and edited an additional book.

10. I have extensive experience in the areas of search engines, machine learning, document classification and information retrieval, image classification,

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

deep learning, and explainable artificial intelligence (“AI”), including: (a) creating and teaching a graduate course at UC Irvine entitled Information Retrieval, Filtering, and Classification (ICS 207); (b) teaching tutorials on document classification, machine learning and search at conferences put on by the International Joint Conferences on Artificial Intelligence (“IJCAI”), the Association for the Advancement of Artificial Intelligence (“AAAI”), and the Association for Computer Machinery (“ACM”) (specifically, ACM’s conferences on Intelligence User Interfaces (“IUI”)); (c) running a company that handled news, classified, and restaurant searches for the Los Angeles Times and San Diego Union Tribune websites; (d) directing the Information and Intelligent Systems Division at the NSF, which provided funding for digital libraries, internet search, databases and machine research under my tenure; (e) receiving an AAAI Classic Paper award on personalization in web search; (f) creating a specialized search engine that classified proposals for the NSF to find reviewers for research proposals; (g) training and advising multiple PhD graduates who have worked on commercial search engines at Yahoo, Google, and Yandex; (h) serving on the Technical Advisory Board on Internet Search for Microsoft; and (i) applying deep learning methods to medical images for the purposes of detecting diseases such as glaucoma and pneumonia.

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

11. I received many honors, awards, research grants and gifts. In 2014, I was awarded the AAAI Classic Paper award for a 1996 publication “for significant contributions to the field of personalizing Internet content and learning user profiles.” *AAAI Classic Paper Award*, AAAI, <https://aaai.org/about-aaai/aaai-awards/aaai-classic-paper-award/> (last visited Aug. 25, 2025). In 2003, I was recognized as a AAAI Fellow “for significant contributions to the development and understanding of machine learning algorithms and their relationship to human learning.” *Elected AAAI Fellow*, AAAI, <https://aaai.org/about-aaai/aaai-awards/the-aaai-fellows-program/elected-aaai-fellows/#2003>. I have received 37 grants worth roughly \$16.9 million in total. From 2019 to 2021, I served on the Defense Science Board of the U.S. Department of Defense (“DOD”), which provides the Secretary of Defense with independent advice and recommendations on scientific, technical, and other matters of special interest to the DOD.

12. My academic credentials include two degrees from the University of Connecticut: (1) a Bachelor of Science degree in Computer Engineering earned in 1980; and (2) a Master of Science degree in Computer Engineering earned in 1981. I earned a Doctor of Philosophy (Ph.D.) in Computer Science from the University of California, Los Angeles, in 1988.

III. INFORMATION CONSIDERED

13. In preparation for this declaration, I have considered the materials discussed in this declaration, including, for example, the '120 Patent, the references cited by the '120 Patent, the prosecution histories of the '120 Patent and applications from which it derives (including the references cited therein), various background articles and materials referenced in this declaration, and the prior art references identified in this declaration. In addition, my opinions are further based on my education, training, experience, and knowledge in the relevant field.

IV. RELEVANT LEGAL STANDARDS

14. I am not an attorney and offer no legal opinions. For the purposes of this Declaration, I have been informed about certain aspects of the law that are relevant to my analysis, as summarized below.

A. Claim Interpretation

15. I have been informed that during an *inter partes* review proceeding, claims are to be construed in light of the specification as would be read by a person of ordinary skill in the relevant art at the time the application was filed. I have been informed that claim terms are given their ordinary and customary meaning as would be understood by a person of ordinary skill in the relevant art in the context of the entire disclosure. A claim term, however, will not receive its ordinary meaning if the patentee acted as his own lexicographer and clearly set forth a definition of the claim

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

term in the specification. In that case, the claim term will receive the definition set forth in the patent.

16. I have been informed that certain patent claim terms may be interpreted as “means-plus-function” claim terms. I have been informed that terminology is part of the U.S. Patent Law in 35 USC §112, paragraph 6 (pre-AIA) / 35 USC § 112(f) (post-AIA), which states the following: “An element in a claim for a combination may be expressed as a means or step for performing a specified function without the recital of structure, material, or acts in support thereof, and such claim shall be construed to cover the corresponding structure, material, or acts described in the specification and equivalents thereof.”

B. Perspective of One of Ordinary Skill in the Art

17. I have been informed that a patent is to be understood from the perspective of a hypothetical person of ordinary skill in the art (“POSITA”). Such an individual is considered to possess normal skills and knowledge in a particular technical field (as opposed to being a genius). I have been informed that in considering what the claims of a patent require, what was known prior to that patent, what a prior art reference discloses, and whether an invention is obvious or not, one must use the perspective of such a POSITA.

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

C. Anticipation

18. I have been informed that a claim is not patentable if it is anticipated. I have been informed that anticipation of a claim requires that every element of a claim is disclosed expressly or inherently in a prior art reference, arranged as in the claim, when considered from the perspective of a person of ordinary skill in the relevant art. I have been informed that when the structure recited in a reference is substantially identical to that of the claims, claimed properties or functions are presumed to be inherent. I have been informed that extrinsic evidence may be used to explain but not expand the meaning of terms and phrases used in the reference relied upon as anticipatory of the claimed subject matter.

D. Obviousness

19. I have been informed that a patent claim is obvious under 35 U.S.C. §103, and therefore invalid, if the claimed subject matter, as a whole, would have been obvious to a POSITA as of the priority date of the patent based on one or more prior art references and/or the knowledge of a POSITA.

20. I have been informed that an obviousness analysis must consider (1) the scope and content of the prior art, (2) the differences between the claims and the prior art, (3) the level of ordinary skill in the pertinent art, and (4) secondary considerations, if any, of non-obviousness (such as unexpected results, commercial

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

success, long- felt but unmet need, failure of others, copying by others, and skepticism of experts).

21. I have been informed that a prior art reference may be combined with other references to disclose each element of the invention under 35 U.S.C. § 103. I have been informed that a reference may also be combined with the knowledge of a POSITA, and that this knowledge may be used to combine multiple references. I have been informed that a POSITA is presumed to know the relevant prior art. I have been informed that the obviousness analysis may take into account the inferences and creative steps that a POSITA would employ.

22. In determining whether a prior art reference would have been combined with other prior art or other information known to a POSITA, I have been informed that the following principles may be considered:

- whether the references to be combined involve non-analogous art;
- whether the references to be combined are in different fields of endeavor than the alleged invention in the Patent;
- whether the references to be combined are reasonably pertinent to the problems to which the inventions of the Patent are directed;
- whether the combination is of familiar elements according to known methods that yields predictable results;
- whether a combination involves the substitution of one known element for another that yields predictable results;

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

- whether the combination involves the use of a known technique to improve similar items or methods in the same way that yields predictable results;
- whether the combination involves the application of a known technique to a prior art reference that is ready for improvement, to yield predictable results;
- whether the combination is “obvious to try”;
- whether the combination involves the known work in one field of endeavor prompting variations of it for use in either the same field or a different one based on design incentives or other market forces, where the variations are predictable to a POSITA;
- whether there is some teaching, suggestion, or motivation in the prior art that would have led one of ordinary skill in the art to modify the prior art reference or to combine prior art reference teachings to arrive at the claimed invention;
- whether the combination requires modifications that render the prior art unsatisfactory for its intended use;
- whether the combination requires modifications that change the principle of operation of the reference;
- whether the combination is reasonably expected to be a success; and
- whether the combination possesses the requisite degree of predictability at the time the invention was made.

23. I have been informed that in determining whether a combination of prior art references renders a claim obvious, it is helpful to consider whether there is some teaching, suggestion, or motivation to combine the references and a reasonable

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

expectation of success in doing so. I have been informed, however, that a teaching, suggestion, or motivation to combine is not required.

V. SUMMARY OF OPINIONS

24. It is my opinion that the Challenged Claims of the '120 Patent are unpatentable under 35 U.S.C. §103 on the following grounds:

Ground	Reference(s)	Basis	Claims
1	Anticipated By U.S. Patent 6,004,015 (“Watanabe”) (EX1006), and/or Obvious Over Watanabe (EX1006), and/or Obvious Over Watanabe (EX1006) in View of F. Friedrichs, <i>et al.</i> , “Evolutionary tuning of multiple SVM parameters,” pp107-117 (January 7, 2005) (“Friedrichs”) (EX1011).	§102/§103	1-8, 11-16
2	Obvious Over Watanabe (EX1006) and Friedrichs (EX1011), in view of Pieter-Tjerk De de Boer <i>et al.</i> , “A tutorial on the Cross-Entropy Method,” ANNALS OF OPERATIONS RESEARCH 134:19-67 (2005) (“de Boer”) (EX1007).	§103	9
3	Obvious Over Watanabe (EX1006) and Friedrichs (EX1011), in view of de Boer (EX1007), M. Kirsten <i>et al.</i> , “Distance Based Approaches to Relational Learning and Clustering,” in Relational Data Mining, Chpt. 9, pp. 213-232 (2001) (“Kirsten”) (EX1008), and Xu <i>et al.</i> "Survey of clustering algorithms." IEEE Transactions on neural networks 16, no. 3 (2005): 645-678. (“Xu”) (EX1017).	§103	10

VI. TECHNOLOGY BACKGROUND

25. The '120 Patent relates to optimizing parameters (referred to therein as “hyperparameters”) of machine learning “classifiers” used to classify objects (e.g., to assist fruit picking robots). EX1001, 1:6-28.

26. The '120 Patent does not define the term “hyperparameter” but indicates that it is a parameter within a set of values (i.e., a solution vector) associated with an “extremely large size” search space wherein “the optimal solution [vector] typically is a rare event.” EX1001, 1:52-55.

27. The '120 Patent claims an optimization adjusting method for such a solution vector (e.g., an optimized hyperparameter vector). EX1001, 2:1-4. The method includes “[d]etermining *optimal* sets of hyperparameters” for an “extremely large size ... *search space*” (EX1001, 1:52-54 (emphasis added)) by “*iteratively* producing an *estimate*” of a target hyperparameter vector. EX1001, 2:1-5 (emphasis added). The claimed process “*iteratively*” produces “an *estimate* of a target hyperparameter vector” using “a hyperparameter vector producing the *best result* in the present and any previous iterations”. EX1001, Claim 1 (emphasis added).

28. However, “optimum adjustment of solution vectors” for parameters associated with an extremely large search space of a machine learning classification

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

process was well known in the prior art, as will be described herein. (*See, e.g.*, EX1006, Abstract.)

29. The '120 Patent discloses that “[h]yperparameters may be determined using various methods” including “statistical optimisation methods” and are expressed as “vectors”. EX1001, 1:29-30, 2:1-5, 2:34-37, Claim 1.

30. The '120 Patent states that not all statistical methods are suitable for determining hyperparameters (EX1001, 1:29-33) and describes problems associated with “the extremely large size of the *search space*” (*id.*, 1:52-54) (emphasis added).

31. The '120 Patent addresses the alleged unsolved problem of a large size search space “by *iteratively* producing an *estimate*” of a target hyperparameter vector. EX1001, 2:1-5 (emphasis added). *See also id.*, Abstract (“iteratively produces an estimate”), 7:15-18 (claim 1).

32. The Challenged Claims recite determining hyperparameters of a classifier in a machine learning system by iteratively producing an estimate of a target hyperparameter vector. *See* EX1001, 7:14-8:66.

33. The iterative steps include drawing a random sample of hyperparameter vectors from a set of possible hyperparameter vectors, updating the estimate of the target hyperparameter vector by using the random sample, and selecting, from the

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

random sample of hyperparameter vectors, a hyperparameter vector producing a best result in the present and any previous iterations. *See* EX1001, 7:14-8-66.

34. The '120 Patent's "Background" and prosecution history do not refer to the prior art's recognition of the exact same search space problem or known statistical methods for optimizing solution vectors to solve that very issue. (Watanabe discloses "an *optimization* adjusting method and an optimization adjusting apparatus capable of *effectively conducting searches* for optimal solutions". EX1006, 1:6-8 (emphasis added). Watanabe recognizes that earlier techniques "take[] a longer time *to search* for solutions" and "oscillat[e] through the *search space*" *Id.*, 5:1-8 (emphasis added).)

35. The term hyperparameter as characterized in the '120 Patent conflates model parameters and configuration parameters of a decision function.

36. The '120 Patent states that "hyperparameters HP determine the decision function of the classifier that classifies items." EX1001, 4:49-50. The '120 Patent's classification process is "controlled by hyperparameters which determine, for example, decision criteria such as threshold values." EX1001, 1:13-23. "Hyperparameters may be determined using various methods, for example heuristic or statistical optimisation methods." EX1001, 1:29-30.

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

37. Classification involves machine learning models defined by parameters that include such “hyperparameters”.

38. The foregoing statements use the term “hyperparameter” to describe parameters of a classification process, and thereby conflate the terms “hyperparameters” and “parameters” in such a way that “hyperparameters” of the ‘120 Patent include the model parameters.¹

39. To place the ‘120 Patent’s use of the term “hyperparameters” in context, a POSITA at the time of the ‘120 Patent’s filing would have recognized and understood that classifiers had been developed in machine learning with neural networks.

40. Neural networks of the ‘120 Patent and earlier used customized training algorithms to increase the speed of learning or the accuracy of the learned classifier.

¹ Model parameters (e.g., connection weights of a model) control the classification process, whereas parameters which have been more recently characterized as “hyperparameters” control the classification learning process (e.g., the learning rate of a genetic algorithm, where the learning rate is set in advance of using the genetic algorithm to train/learn the model parameters (e.g., connection weights) that will be used by the model).

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

41. The term “*hyperparameter*” was not prevalent at the time of the ’120 Patent but had been used previously in the statistical literature to denote parameters for fitting a model to data (i.e., training), for defining structure and behavior of training algorithms, etc., and had been recognized as being distinct from the model parameters.

42. In the ’120 Patent, hyperparameters can directly affect model accuracy and convergence. EX1001, 2:31-32; 6:18-20.

43. Hyperparameters can be thresholds associated with a given training model (i.e., thresholds that are decisions made before the learning process begins).² EX1001, 1:20-24.

44. The iterative hyperparameter optimization disclosed by the ’120 Patent was well-known and described in prior art before the ’120 Patent was filed.

VII. THE CHALLENGED ’120 Patent

45. The ’120 Patent contains 16 claims. I have been asked to opine on the validity of Claims 1-16.

²Hyperparameters can be adjusted during the training process, but the criteria for the change (e.g., the learning rate) are determined before the training process.

VIII. THE '120 PATENT PROSECUTION HISTORY

46. I have reviewed the file history of the '120 Patent. The application which issued as the '120 Patent was allowed in the first office action without any rejection. EX1002, pp.15-28.

47. Amendments filed October 23, 2009 and July 8, 2010 (*see* EX1002, pp.79-85; pp.49-54) only addressed formality issues and clarity issues raised by the EPO during European examination of counterpart EP Application 08741668.1³ (EX1009).

48. An Examiner's Amendment further addressed formality issues on January 11, 2013 after which the '120 Patent was granted. *See* EX1002, pp.15-28.

49. The counterpart EP application was abandoned and never granted as an EP patent. EX1009, pp.1-3.

IX. LEVEL OF ORDINARY SKILL IN THE ART

50. A person of ordinary skill in the art ("POSITA") pertinent to the '120 Patent would have had at least a Master's degree in electrical engineering, computer science, or a related discipline, knowledge of machine learning classification

³EP 08741668.1, filed April 25, 2008, was abandoned after lack of novelty and inventive step rejections. EX1009.

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

techniques as of this time, and at least two years of experience working in a field involving classifier development in machine learning systems. A person with a different degree but with additional relevant experience could still qualify if the additional experience compensates for the different educational background. A person with less than two years of experience working in a field involving networking and system security could still qualify if the person has additional education that compensates for their lesser experience.

X. PRIORITY DATE

51. I have been informed that the “priority date” or “earliest effective filing date” of a patent is the date on which it is filed, or the date on which an earlier-filed U.S. or international patent application was filed if the patentee claims the benefit of priority to that earlier-filed U.S. or international patent application.

52. The '120 Patent issued from U.S. Patent Application No. 12/597,257, filed on July 8, 2010, which is a U.S. National Stage of PCT/NL08/50247, filed on April 25, 2008 (published as WO 2008/133509) (EX1003), and claims priority to EP 07112037.2, filed on July 9, 2007 (EX1004) and EP 07106963.7, filed on April 25, 2007 (EX1005). For purposes of this declaration, I have assumed that the '120 Patent is entitled to a priority date of April 25, 2007, and I have not evaluated or

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

formed an opinion on whether the claims of the '120 Patent are actually entitled to that date.

XI. CLAIM CONSTRUCTION

53. For purposes of my analysis in this IPR proceeding, I have been informed that terms that appear in the claims of the '120 Patent should be interpreted according to their plain and ordinary meaning under the *Phillips* standard. In determining the ordinary and customary meaning, I have been informed that the words of a claim are first given their plain meaning that those words would have had to a POSITA at the time of the alleged invention. I also have been informed that the claims, specification, and file history may be used to better construe a claim insofar as the plain meaning of the claims cannot be understood. I have been informed that even treatises and dictionaries may be used under limited circumstances to determine the meaning attributed by a POSITA to a claim term at the time of filing.

54. I also understand that, for terms that evoke a means plus function construction, “the construction of the claim must identify the specific portions of the specification that describe the structure, material, or acts corresponding to each claimed function.” 37 C.F.R. §42.104(b)(3).

55. I understand from counsel that construing a means-plus-function claim term is a two-step process that includes (1) identifying the claimed function and (2)

then determining what structure, if any, disclosed in the specification corresponds to the claimed function.

XII. OVERVIEW OF THE PRIOR ART REFERENCES

A. Watanabe (EX1006)

56. Watanabe, which was not cited in the '120 Patent prosecution history, addresses the exact problem of the '120 Patent by determining “optimum adjustment of solution vectors” for parameters associated with an extremely large search space of a machine learning classification process. EX1006, Abstract; *see also*, EX1006 in its entirety.

57. Watanabe discloses parameters within a set of values (i.e., a solution vector) associated with an extremely large-sized search space, and further discloses an iterative optimization method to find an optimal solution vector within that space. EX1001, 1:52-55.

58. Watanabe optimizes a “vector solution” for machine learning classification using an “*optimization* adjusting method” for “*effectively conducting searches for optimal solutions*”. *See, e.g.*, EX1006, 1:6-8 (emphasis added); *see also id.*, 5:2-8 (addressing known problems of a large search space, such as a “longer time to search” and “oscillating through the search space” to converge on a solution).

59. Watanabe discloses the exact “solution vector” optimization process of the '120 Patent for parameters characterized in the '120 Patent as “hyperparameters”.

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

60. Watanabe teaches to “***iteratively***” produce an “***estimate***” of a target vector using “a solution vector having the ***highest fitness*** in the most recent solution vector set” and “successively updating” the solution vector set. *See, e.g.*, EX1006, 5:53-64. “By iteratively performing this process, an optimal solution is estimated” as a vector of hyperparameters as defined by the ‘120 Patent. *Id.*, at 3:3-11, 22:21-31, 23:18-37, 30:14-18.

61. Watanabe optimizes “genetic algorithms for optimal solution estimation problems.” EX1006, 1:6-15.

62. Although Watanabe does not expressly use the term “hyperparameters”, Watanabe’s iterative optimization method is applied to optimize parameters within a solution vector containing “hyperparameters” as that term is used in the ’120 Patent. *See* EX1006, 47:20-24. For example, Watanabe discloses optimizing “connection weights” as a solution vector of a neural network model.

63. One skilled in the art would have considered it appropriate to optimize that solution vector using Watanabe’s disclosed genetic optimization algorithms.

64. Alternately, the connection weights would have been learned using Watanabe’s disclosed back propagation learning method configured with an optimized solution vector of configuration parameters (i.e., hyperparameters α and η).

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

65. One skilled in the art would have considered it appropriate to optimize that solution vector using Watanabe's disclosed genetic optimization algorithm. EX1006, 49:25-26. .

66. The "back propagation" learning method is applied to learning the "connection weights" using Mathematical Expression 21 of Column 49. *See* EX1006, 49:1-28.

67. Watanabe's disclosed solution vector optimization method would have been considered appropriate because Watanabe teaches applying its optimization method for "effectively conducting searches for optimal solutions." EX1006, 1:6-8; *see also* 5:2-8.

68. Watanabe's parameter optimization solution iteratively produces an estimate of a target parameter vector to effectively search a large size search space. EX1006, 5:53-64.

69. Watanabe discloses "***successively updating*** the solution vector set and thereby ***estimating*** a nearly optimal solution vector for the optimization problem being solved"(*Id.*, 3:9-11), and discloses that "a solution vector having the highest fitness in the most recent solution vector set is output as the optimal solution vector." (*Id.*, 30:18-18, Fig. 3, "Initialization of Iteration Count 1").

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

70. In Watanabe's iterative vector optimization process, "an information presenting section presents information expressed by a solution vector"; a "model estimation executing section generates an evaluation model for an adjustment process"; a "fitness calculating section calculates fitness from the evaluation made"; and "in accordance with the thus calculated fitness, a recombination operation section performs an arithmetical recombination operation." EX1006, Abstract.

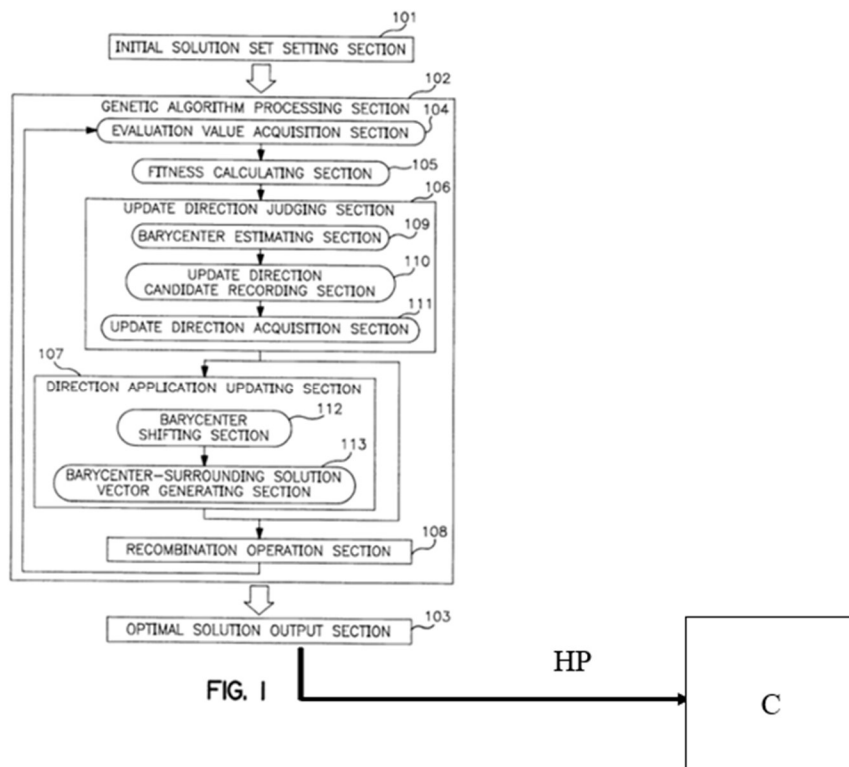
71. In Watanabe's process, "solution vectors of low fitness are eliminated from the current solution set, and solution vectors of high fitness are selected for survival" to generate an "estimation of an optimal solution vector." EX1006, 1:65-2:1. "The estimation of an optimal solution vector is performed by iteratively performing the above process until either of the termination conditions, 1 or 2, is satisfied." EX1006, 27:44-46.

72. In Watanabe's FIG. 1, section 101 sets "an initial set of solutions, $P=\{p_k\}(k=1, \dots, n)$ "; 102 is a genetic algorithm processing section for performing the estimation of an optimal solution by genetic recombination operations, and 103 is an optimal solution output section for outputting as the optimal solution vector a solution vector having the highest fitness in the most recent solution vector set when a predetermined termination condition is satisfied." EX1006, 22:21-31.

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

73. Watanabe provides classification processes in neural networks. *See, e.g.* EX1006, 47:6-49:52. Genetic algorithms and genetic-based classifier tuning algorithms utilizing crossover, mutation, recombination, elitism (*see* EX1006, 67:30-35) such as those described by Watanabe were well known.

74. FIG. 1 (annotated below) shows an iterative, genetic solution vector process to optimize Watanabe's disclosed learning of connection weights within a neural network, wherein the solution vector of connection weights (annotated as "HP") would have been optimized for a machine learning classification process (annotated as "C") (*see, e.g.*, references to "Machine Learning" and "highly intelligent pattern classification" at EX1006, 1:34-40; 51:7-18):



Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

75. Watanabe's column 49 "Mathematical Expression 21" is associated with alternative algorithms for adjusting weights (*see id.*, 49:36-51), and with alternative training algorithms (*see id.*, 51:7-18), all of which involve an optimized solution vector.

76. Watanabe recognizes the need/benefit of determining optimal parameters for any solution vector.

77. Watanabe recognizes the use of hyperparameters in evaluating/adjusting training algorithms. Watanabe states that "an evaluation model for the user adjustment process can be estimated based on the history of adjustments made by the user when adjusting the solution vector optimum for his requirements, and furthermore, using this evaluation model, it is possible to evaluate a given solution vector." EX1006, 49:34-39.

78. Watanabe describes alternative learning mechanisms (e.g., back propagation, quasi-Newton, and learning vector quantization) and how the learning back propagation training method can be controlled by various parameters for adjusting weights. *See id.*, 49:11-51. Watanabe, at 49:25-26, discusses a change to connection weights α and η , which are the hyperparameters of a hyperparameter vector.

79. A POSITA would have understood Watanabe's evaluating/adjusting training algorithms to be configured with a solution vector of optimized hyperparameters.

B. Friedrichs (EX1011)

80. Friedrichs discloses applying an automated iterative, genetic (i.e., "evolutionary") optimization method to optimize a "hyperparameter" vector:

"The problem of model selection for support vector machines (SVMs) is considered. We propose an evolutionary approach to determine multiple SVM hyperparameters."

EX1011, Abstract, p. 107.

81. Just as the '120 Patent purports to replace a "simple grid search" with an automated iterative, evolutionary optimization process that invokes "elitism" (*see* EX1006, 67:30-35) to determine a hyperparameter vector (EX1001, 1:47-48), Friedrichs recognizes the limitations of determining hyperparameters via grid search (*see* EX1011, 108:8-12) and "propose[s] an evolutionary method for hyperparameter selection" (EX1011, ln. 27, p. 108). *See also* EX1011 in its entirety.

82. "**Evolutionary algorithms** have been successfully applied to model selection for neural networks. This includes the recent applications of **genetic algorithms** for feature selection of SVMs." (EX1011, 108:28-30. Friedrichs, like

Watanabe, therefore discloses optimizing genetic algorithms for optimal solution estimation problems *See, Id.*; EX1006, 1:6-15.

C. de Boer (EX1007)

83. de Boer discloses updating estimates of a target hyperparameter vector “v” using an optimization-based, formulaic approach for “applications in combinatorial optimization and machine learning.” EX1007, Abstract; *see also* EX1007 in its entirety.

1. Define $\hat{v}_0 := u$. Set $t := 1$ (iteration counter).
2. Generate a random sample $X_1, \dots, X_N$ according to the pdf $f(\cdot; \hat{v}_{t-1})$. Calculate the performances $S(X_i)$ for all i , and order them from smallest to biggest, $S_{(1)} \leq \dots \leq S_{(N)}$. Let $\hat{\gamma}_t$ be the $(1 - \rho)$ sample quantile of performances: $\hat{\gamma}_t := S_{(\lceil (1-\rho)N \rceil)}$, provided this is less than γ . Otherwise, put $\hat{\gamma}_t := \gamma$.
3. Use the same sample to calculate, for $j = 1, \dots, n (= 5)$,

$$\hat{v}_{t,j} = \frac{\sum_{i=1}^N I_{\{S(X_i) \geq \hat{\gamma}_t\}} W(X_i; u, \hat{v}_{t-1}) X_{ij}}{\sum_{i=1}^N I_{\{S(X_i) \geq \hat{\gamma}_t\}} W(X_i; u, \hat{v}_{t-1})}.$$

EX1007, p 23.

84. de Boer’s formula applies a multi-purpose cross-entropy (CE) method with applications in “neural computation.” EX1007, p. 21.

85. The ’120 Patent states that “[t]he present invention allows a method similar to the cross-entropy (CE) method to be used for estimating and determining hyperparameters.” EX1001, 4:61-63.

86. A POSITA would have recognized the usefulness of de Boer’s CE formula in determining the parameters for training a classifier, i.e., hyperparameters. de Boer describes this formula as being a “general CE algorithm”.

D. Kirsten (EX1008)

87. Kirsten discloses updating estimates of a target hyperparameter vector “v” using an optimization-based, formulaic approach that includes a weighting function of a distance-based approach to relational learning in clustering. *See* EX1008, Abstract; *see also* EX1008 in its entirety.

88. A first-order instance-based learner applying a classical k-nearest neighbor classifier is discussed in which a weighted average is used to determine how much of an influence should be allocated to farther-away neighbors by using:

$$h(x_{new}) := \frac{\sum_{e=(x,y) \in N} sim(x, x_{new})^a \cdot y}{\sum_{e=(x,y) \in N} sim(x, x_{new})^a}. \text{ See EX1008, pp. 220-221.}$$

89. Kristen’s formula is similar to the formula recited by Claim 9 of the ‘120 Patent, and uses a similarity function $sim(x, y) := 1 - dist(x, y)$ so that the “distance measure takes on values in the range [0,1]”. EX1008, p. 220.

90. The ‘120 Patent states “that when the Euclidean distance between E^t and X^t_{ij} is zero, the weighting function W is equal to one.” EX1001, 3:26-27.

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

91. The similarity function of Kristen performs the same role as the weighting function in Claim 10 of the '120 Patent, as each of the similarity function and the weighted function places an emphasis (i.e., gives higher weight) on vectors closest to a target vector.

92. The only operational constraint of significance in the '120 Patent's weighting function is for "W [to be] a weighting function and $I\{S(X_i^t) \geq \gamma^t\}$ [to be] an indicator function which is equal to 1 if $S(X_i^t)$ greater than or equal to the threshold value γ^t , and else is equal to 0". EX1001, 3:11-15. Kristen's similarity function performs this same role.

E. Xu (EX1017)

93. Xu discloses a Cosine Similarity function, a Euclidean distance function and a city block distance function, and demonstrates that these similarity functions for defining distance and similarity measures were commonly used in classical machine learning before the filing date of the '120 Patent. EX1017, Table 1, p. 648; *see also* EX1017 in its entirety. The city block function is also referred to as the Manhattan distance function.

**XIII. DETAILED EXPLANATION OF THE UNPATENTABILITY
GROUNDS**

94. In my opinion, the Challenged Claims are unpatentable over the prior art.

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

**A. Claims 1-8 and 11-16 Are Anticipated by Watanabe (EX1006),
And/Or Would Have Been Obvious Over Watanabe (EX1006),
And/Or Would Have Been Obvious Over Watanabe in View of
Friedrichs (Ground 1)**

95. The following sections reference where elements of Claims 1-8 and 11-16 are anticipated by/obvious over the prior art.

**1. Claim 1 (Preamble): “A method of determining
hyperparameters of a classifier in a machine learning
system by iteratively producing an estimate of a target
hyperparameter vector, each iteration comprising the steps
of:”**

96. To the extent the Preamble is limiting, Watanabe discloses and suggests the Preamble.

97. Watanabe discloses a method (“optimization adjusting method” (EX1006, 1:6-8)) of determining “solution vectors” (*id.*, Abstract; 1:16-19) that contain parameters used for classification (i.e., in a classifier) of a machine learning system. EX1006, 51:14 (referencing “highly intelligent pattern classification”). *See also* EX1006 in its entirety.

98. Although Watanabe does not use the term “hyperparameters”, the parameters within a solution vector of Watanabe constitute “hyperparameters” as described by the ’120 Patent.

99. Watanabe addresses the exact problem of the ’120 Patent by determining “optimum adjustment of solution vectors” for parameters associated

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

with an extremely large search space of a machine learning classification process.

EX1006, Abstract.

100. Watanabe discloses optimizing a “vector solution” for machine learning classification using an “**optimization** adjusting method” for “**effectively conducting searches** for **optimal** solutions”. *See, e.g.*, EX1006, 1:6-8 (emphasis added); *see also id.*, 5:2-8 (addressing known problems of a large search space, such as a “longer time to search” and “oscillating through the search space” to converge on a solution).

**i) Watanabe Expressly Discloses Optimizing a Solution Vector
of Parameters that Correspond to “Hyperparameters”**

101. Watanabe discloses the exact “solution vector” optimization process of the ‘120 Patent for parameters characterized in the ‘120 Patent as “hyperparameters”.

102. Watanabe teaches to “**iteratively**” produce an “**estimate**” of a target vector using “a solution vector having the **highest fitness** in the most recent solution vector set” and “successively updating” the solution vector set. *See, e.g.*, EX1006, 5:53-64. “By iteratively performing this process, an optimal solution is estimated” as a vector of hyperparameters as defined by the ‘120 Patent. *Id.*, at 3:3-11, 22:21-31, 23:18-37, 30:14-18.

103. Watanabe discloses optimizing known “genetic algorithms for optimal solution estimation problems.” EX1006, 1:6-15. Although Watanabe does not

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

expressly use the term “hyperparameters”, Watanabe’s iterative optimization method is applied to optimize parameters within a solution vector containing “hyperparameters” as that term is used in the ’120 Patent. *See* EX1006, 47:20-24.

104. Watanabe discloses optimizing “connection weights” of a connection weight solution vector within a neural network model. One skilled in the art would have considered it appropriate to optimize that solution vector using Watanabe’s disclosed genetic optimization algorithms. The optimized solution vector of connection weights constitutes a “hyperparameter vector” as that term is described in the ’120 Patent. *See* U.S. Patent No. 5,249,259 (“Harvey”) (EX1025), which demonstrates the knowledge of those skilled in the art would have understood Watanabe’s genetic algorithms to be used for optimizing Watanabe’s disclosed connection weights. *Id.*, at Abstract:1-3.

105. Accordingly, Watanabe discloses “determining hyperparameters” of a hyperparameter vector as recited in Claim 1.

**ii) Watanabe Discloses Optimizing Any Solution Vector, and
Therefore Teaches/Suggests Optimizing the Disclosed
Solution Vector Containing Configuration Parameters (e.g.,
 α and η) of The Back Propagation Learning Method**

106. To the extent the learned “connection weights” are not themselves considered “hyperparameters” by reference to Watanabe alone, it was well recognized that back propagation learning methods involved numerous

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

configuration parameters (i.e., hyperparameters), such as model capacity/batch size (number of connection weights), “early stopping” to avoid overfitting of the model to the batch, and learning rate of the model. (See, added details “back propagation” learning method as disclosed by Caruana (EX1016).

107. Watanabe’s disclosed back propagation learning method (e.g., as applied to learning of model connection weights) would have been configured with an optimized solution vector of configuration parameters (i.e., α and η “hyperparameters”). One skilled in the art would have considered it appropriate to optimize such a solution vector using Watanabe’s disclosed genetic optimization algorithm. EX1006, 49:25-26.

108. A POSITA reading Watanabe’s disclosed solution vector optimization method would have applied that solution to the back propagation configuration parameters α and η because Watanabe teaches applying its optimization method for “effectively conducting searches for optimal solutions.” EX1006, 1:6-8; see also 5:2-8. Watanabe teaches applying its optimization method for “effectively conducting searches for optimal solutions.” EX1006, 1:16-821-24.

109. Watanabe discloses that the back propagation method includes configuration parameters α and η (i.e., “hyperparameters”). EX1006, 49:25-26. Such

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

a solution vector containing α and/or η would have been “constructed appropriately” using the optimization methods disclosed by Watanabe. (*See* EX1006, 49:1-28).

110. A POSITA would have understood the applicability of Watanabe’s disclosed genetic “optimization adjusting method” to determining “solution vectors”. A POSITA would have understood Watanabe’s discussion of “estimating a nearly optimal solution vector for *the optimization problem being solved*”, (EX1006, Abstract, 1:6-19, 3:9-11, 49:1-28), to be an optimization method for determining hyperparameters.

111. Such an application of Watanabe’s optimization process to a hyperparameter vector would have been predictable, and performed in predictable fashion with a high probability of success, as an iterative technique applied for optimizing any parameter.

112. Watanabe’s “solution vectors” containing hyperparameters are optimized via optimization adjusting methods and a genetic algorithm (*Ibid.*; *see also* 22:6-20).

113. Watanabe discloses “model estimation executing section 3804 performs the training of this neural network” (*id.*, 48:66-67)) to determine the parameters of the solution vectors. Watanabe’s disclosure of back propagation involves propagation model selection/configuration.

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

114. A “back propagation” training algorithm adjusts weights (*id.*, 49:11-16), parameters for “adjusting the solution vector optimum for his requirements” (*id.*, 49:34-39), and parameters for selecting between “model estimation based on the pattern classification by the back-propagation... [and] a neural network using a learning vector quantization method...” (*id.*, 51:7-18).

115. Watanabe’s disclosed back propagation involves hyperparameters known to be used in back propagation model selection/configuration, as further evidenced by Caruana (EX1016).

116. Therefore, it would have been obvious to a POSITA to use Watanabe’s disclosed “optimization adjusting method” of determining “solution vectors” for optimizing the hyperparameters of Watanabe’s back propagation learning method.

**iii) A POSITA Would Have Been Motivated to Use Watanabe’s
Genetic Optimization Method to Optimize the Solution
Vector Containing Parameters α and η Based on Friedrichs’s
Teachings Regarding Genetic “Hyperparameter”
Optimization**

117. Although Watanabe does not expressly refer to “hyperparameters”, Watanabe’s express disclosure of using genetic algorithms to optimize a solution vector, coupled with Watanabe’s disclosure of configuration parameters α and η , discloses/suggests use of a genetic optimization method to optimize α and η for reasons expressly disclosed by Friedrichs (i.e., improve upon the limitations of a grid

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

search approach by determining parameters of a hyperparameter solution vector using an evolutionary (e.g., genetic) approach). (*See also* Mannor (EX1010) and Guha (EX1012) regarding the knowledge of those skilled in the art as regards optimizing hyperparameters in solution vectors of machine classification systems.)

118. To the extent Watanabe’s optimization method is not considered to be applied to hyperparameters as described in the ’120 Patent, Friedrichs discloses automating an iterative optimization method to optimize hyperparameters using an evolutionary algorithm – exemplary genetic algorithms used in Watanabe’s optimization method are evolutionary algorithms. *See* EX1011, 108:28-30.

119. The ’120 Patent purports to provide an automated iterative optimization process to improve upon determining hyperparameter values “using a simple grid search” (EX1001, 1:47-48).

120. Friedrichs recognizes the limitations of determining hyperparameters via grid search (*see* EX1011, 108:8-12) and “propose[s] an evolutionary method for hyperparameter selection” (EX1011, 108:27). *See also* EX1011 in its entirety.

121. Watanabe discloses optimizing known “genetic algorithms for optimal solution estimation problems.” (EX1006, 1:6-15). Genetic algorithms are evolutionary algorithms: “Evolutionary algorithms have been successfully applied

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

to model selection for neural networks. This includes the recent applications of genetic algorithms for feature selection of SVMs.” EX1011,108:28-30.

122. Friedrichs demonstrates that an iterative technique applied for optimizing a model parameter (which is expressly disclosed by Watanabe) can be equally applied for optimizing a hyperparameter.

123. To the extent Patent Owner argues that the optimized solution vectors of Watanabe do not constitute an optimized hyperparameter vector, a POSITA would have understood it to be obvious to apply Watanabe’s optimization process to a solution vector of hyperparameters as this is expressly taught by Friedrichs.

124. Friedrichs teaches the benefit of optimizing a solution vector of hyperparameters via an automated iterative optimization method which addresses drawbacks (e.g., “computational complexity”) associated with a grid search. *See* EX1011, 108:8-12.

125. The ’120 Patent purports to provide an automated iterative optimization process to improve upon determining hyperparameter values “using a simple grid search” (EX1001, 1:47-48). Therefore, an application of Watanabe’s optimization process (which is expressly disclosed for providing a solution vector for a model parameter) for determining a hyperparameter would have been predictable, and performed in predictable fashion with a high probability of success.

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

126. Therefore, it would have been obvious to a POSITA to modify Watanabe's disclosed "optimization adjusting method" of determining "solution vectors" for model parameters based on Friedrichs disclosed automated iterative optimization method for determining hyperparameters to provide an optimization adjustment method for determining hyperparameters.

127. Watanabe's method iteratively produces ("iteratively performs" (EX1006, Abstract)) an estimate of a target hyperparameter vector ("optimal solution is estimated" (*id.*, 5:63-64)). *See also id.*, 5:63-64 ("By *iteratively* performing this process, an optimal solution vector is *estimated*") (emphasis added); Fig. 3 ("Initialization of Iteration Count 1").

128. Just as the '120 Patent seeks to determine "*optimal* sets of hyperparameter values" (EX1001, 1:52-54) (emphasis added), Watanabe discloses and suggests "*optimum* adjustment of solution vectors" to obtain "*optimal* solutions". EX1006, Abstract, 1:5-15 (emphasis added).

129. Watanabe discloses "an optimization adjusting method and an optimization adjusting apparatus capable of effectively conducting searches for optimal solutions". EX1006, 1:6-8. In Watanabe's method, "a solution vector having the highest fitness in the most recent solution vector set is output as the optimal

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

solution vector” and Watanabe “performs the estimation of an optimal solution by iteratively performing the above process”. EX1006, 30:14-22.

130. The ’120 Patent discloses that the “hyperparameters may be determined *using various methods*” and are expressed as “*vectors*”. EX1001, 1:29-30, 2:1-5, 2:34-37, claim 1 (emphasis added) (“hyperparameter vectors”). Watanabe discloses determining “solution *vectors*” in a computerized classification process (*i.e.*, a classifier in a neural network) (*see, e.g.* EX1006, 47:6-49:52) (emphasis added), and discloses/suggests determining solution vectors to determine hyperparameters of a classifier (*see, e.g., id.*, 49:11-28 (back propagation); 49:36:51 (alternative algorithms for adjusting weights); 51:7-18 (alternatives for training algorithms)).

131. Petitioner relies on features described by Watanabe across multiple applications/embodiments which all relate to a common iterative process to produce an estimate of a solution vector as an optimal solution. *See* EX1006, 3:3-11, 5:53-64, 22:21-31, 23:18-37, 30:14-18.

132. Watanabe’s iteratively updating a solution vector to find the optimal solution for each application/problem is the same.

133. Watanabe states that “[t]he specific problem treated here and the method of expressing the elements of a solution vector are the same as those described in [other] embodiments of the invention.” *See e.g.*, EX1006, 32:48-51,

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

35:59-62, 37:22-25. Watanabe summarizes the differences between each of its twenty-three embodiments at 5:53-10:5. Each embodiment modifies the initial solution vector set by dividing vector sets into groups, consolidating vector groups into a set, extracting from neighborhood sets to form a vector set, etc., or uses an update region limiting technique to limit the region from which estimates are updated, but continues with the same recombination operation to generate new solution vectors from the current solution vector set and perform the processes iteratively to estimate an optimal solution vector.

134. The components and functionalities of the various embodiments relied upon by Petitioner are compatible and combinable to one another.

135. Any assertion that claimed elements relied upon in Watanabe are not disclosed in a single embodiment of Watanabe is untenable because the relied upon disclosures of Watanabe relate to a common iterative process.

136. Watanabe teaches its embodiments are combinable, such that a POSITA reading Watanabe would have envisaged the claimed method.

137. A POSITA would have recognized that Watanabe discloses the combined use of such components and functionalities, and anticipates a given limitation even if the disclosures relied upon relate to different embodiments.

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

138. A POSITA is presumed to have been motivated to combine elements in a single prior art reference.

139. According to the '120 Patent, hyperparameters may be determined using “statistical methods.” EX1001, 1:29-33.

140. Watanabe discloses statistical methods whereby “solution vectors of low fitness are eliminated from the current solution set, and solution vectors of high fitness are selected for survival” (EX1006, 1:65-67) to generate an “estimation of an optimal solution vector.” (*id.*, 27:44-46.). *See also id.*, 33:2-3 (“vectors ***probabilistically*** having high fitness are selected as representative solution vectors”) (emphasis added). Watanabe’s optimization adjustment method determines optimal sets of values for solution vectors (hyperparameter vectors). Watanabe’s method expressly addresses problems relating to search spaces that the '120 Patent purports to address. *See* EX1001, 1:52-54.

141. Watanabe’s technique facilitates “effectively conducting searches for optimal solutions” (EX1006, 1:6-8) and addresses problems of “longer time to search for solutions” and “oscillat[ing] through the search space” (*Id.*, 4:60—5:8).

142. The '120 Patent classification process is “controlled by hyperparameters which determine, for example, decision criteria such as threshold values.” EX1001, 1:13-23. This statement demonstrates that the '120 Patent

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

conflates model parameters with configuration parameters, and that both can be considered “hyperparameters” for purposes of claim 1.

143. Figure 1 of '120 Patent “schematically shows a classifier having hyperparameters as input control variables,” wherein “[t]he classifier (C) 1 ... receives feature data FD and hyperparameters HP, and produces classification data CD”. EX1001, 4:20-21, 4:29-31. Figure 1 of the '120 Patent is shown below:

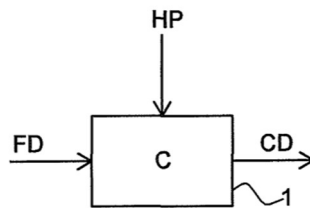
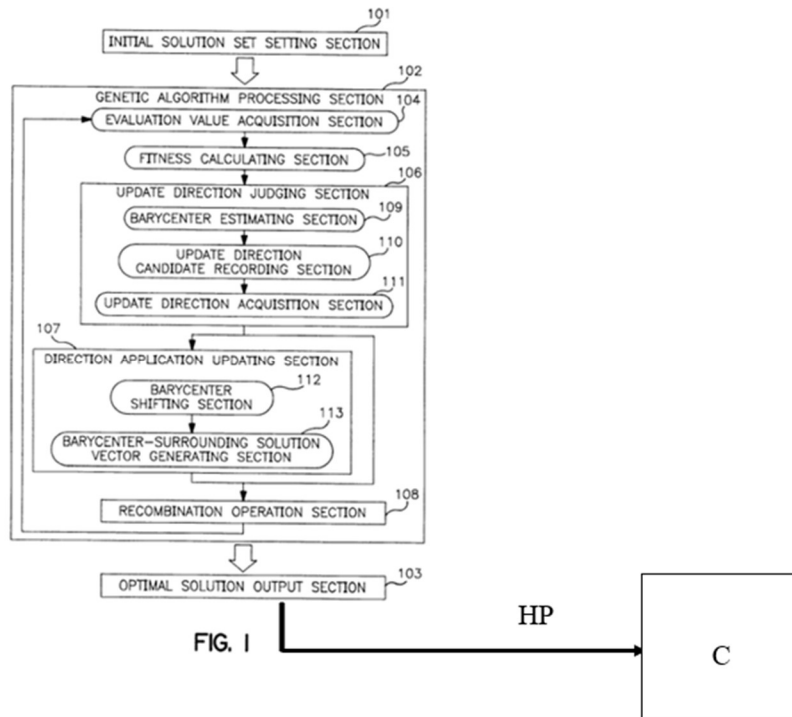


FIG. 1

144. Watanabe’s FIG. 1 (annotated below) shows an iterative solution vector optimization process applicable to Watanabe’s disclosed back propagation learning of connection weights within a neural network, wherein connection weights (annotated as “HP”) are optimized for a neural network of a machine learning classification process (annotated as “C”) (*see, e.g.*, references to “Machine Learning” and “highly intelligent pattern classification” at EX1006, 1:34-40; 51:7-18):

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120



145. Watanabe discloses optimizing solution vectors of parameters. *See, e.g.* EX1006, 47:6-49:52.

146. Watanabe describes alternative learning mechanisms (back propagation, quasi-Newton-and learning vector quantization) and how the learning back propagation training method can be controlled by various parameters for adjusting weights. *See id.*, 49:11-51. Such adjustments in a computerized classification process using Watanabe's disclosed neural network would involve optimizing a hyperparameter for the classifier training.

147. Watanabe therefore discloses optimizing a hyperparameter vector to the same extent as the '120 Patent.

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

148. A POSITA would have understood Watanabe’s optimization method to be applicable to adjusting solution vectors of both model parameters and model configuration parameters—i.e., Watanabe’s solution vectors include hyperparameters regardless how that term is defined.

149. Watanabe’s optimization adjusting method is disclosed as suitable for any intelligent classification process involving a search strategy for a solution vector.

150. This includes optimization of a hyperparameter vector for a classifier in the machine learning environment disclosed by Watanabe. *See also, e.g.,* Friedrichs (EX1011). Neural network classifiers with solution vector parameters that define a neural network’s structure, and solution vectors for tuning classification, were well-known as reflected by the state of art described in Watanabe. EX1006, “BACKGROUND” at 1:5-5:50.

151. Watanabe discloses the “Reasons for Allowance” cited in the ’120 Patent file history which states that “the claimed classifier is inside a special purpose hardware machine learning ‘system’ (i.e., apparatus.)”. EX1002, p.22. The ’120 Patent discloses:

“The classifier may be constituted by an apparatus (hardware embodiment), but is preferably constituted by *a software program* running on a computer apparatus, for example *a general purpose computer (“personal*

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

computer”), a dedicated computer, or other suitable apparatus.”

EX1001, 4:35-40 (emphasis added).

152. A “non-limiting example” of a machine learning system includes a classifier and a device (a processor and associated memory). EX1001, 4:4-14, 4:35-40, 6:46-60.

153. Watanabe similarly discloses a computer-implemented machine-learning system and method. *See, e.g.*, EX1006, Figs. 10, 13, 16 (“genetic algorithm processing section”), Fig. 34 (“main processing section”, “recording medium”), 44:14, 31-33.

154. Watanabe implements its “highly intelligent pattern classification” (EX1006, 51:14) on a “neural network” (*Id.*, at 51:5) which a POSITA would have understood to constitute a classifier running on a special purpose hardware machine learning system. Watanabe specifically discusses an optimization process for determining solution vectors in a “neural network” which a POSITA would have understood to represent a computer system.

155. The Reasons for Allowance also state that “in the art, a hyperparameter is a parameter of a prior distribution; the term is used to distinguish it from parameters of the model for the underlying system under analysis.” EX1002, pp. 22-23. Watanabe discloses and suggests a “hyperparameter” to the same extent as

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

the '120 Patent such that Watanabe's hyperparameters necessarily satisfy the Examiner's characterization of them as parameters of a prior distribution.

156. Regarding the claimed recitation “iteratively producing an estimate of a target hyperparameter vector,” Watanabe discloses this limitation. Watanabe discloses a “selection” operation in which “[s]olution vectors of low fitness are eliminated from the current solution set, and solution vectors of high fitness are selected for survival.” EX1006, 1:65-2:1. “The **estimation** of an **optimal** [i.e., **target**] solution vector is performed by **iteratively** performing the above process until either of the termination conditions, 1 or 2, is satisfied.” EX1006, 27:44-46 (emphasis added). *See also id.*, 30:18-22.

a. [1.1]: “drawing a random sample of hyperparameter vectors from a set of possible hyperparameter vectors,”

157. Watanabe discloses and suggests Element [1.1].

158. Watanabe discloses drawing a “random” sample of hyperparameter vectors (“an initial solution vector set, $P=\{p_k\}(k=1, \dots, n)$... in accordance with a prescribed procedure or by an external instruction”). EX1006, 23:61-64.

159. Watanabe discloses that “[v]arious methods of setting [the initial solution vector set P] are possible” (*id.*, 23:64) and discloses drawing a “random” sample of vectors from a set of possible vectors (*e.g.*, from the possible vectors

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

calculated based on Watanabe's expression method). *Id.*, 23:64, 43:1-4 ("Using uniform random numbers, the initial solution set setting section 101 **randomly selects solution vectors from among the solution vectors** contained within the region Φ , **to construct an initial solution vector set** $P=\{p_k\}$ ($k=1, \dots, n$)), 24:9-20 ("using uniform *random* numbers in $[0,1)$ and using Expression 1, the l -th bit ($l=1, \dots, B_{len} \times m$) of the solution vector expressed in the form of a fixed-length bit string code is obtained, *thereby setting the initial solution vector set* $P=\{p_k\}$ ($k=1, \dots, n$).") (emphasis added).

160. The sample of vectors drawn in Watanabe is a "random sample" as claimed because Watanabe's sample is completely random given Watanabe's use "random numbers" to set the initial solution vector set P . Therefore, Watanabe's sample of vectors constitutes a "pseudo-random sample" (*see* EX1001, 2:40-43) contemplated by the '120 Patent for the same reason.

**b. [1.2]: "updating the estimate of the target
hyperparameter vector by using the random sample,
and"**

161. Watanabe discloses and suggests Element [1.2].

162. In view of the Preamble, the claimed "estimate of the target hyperparameter vector" refers to the vector that is iteratively produced in the claimed

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

method. EX1001, claim 1 (“by iteratively producing an estimate of a target hyperparameter vector”).

163. Watanabe uses a “recombination operation” to perform the updating procedure prescribed by Element [1.2].

164. Watanabe starts with “an **initial set of solutions**, $P=\{p_k\}$ ($k=1, \dots, n$) **[i.e., the random sample – discussed *supra*], which is subsequently *updated*...**” EX1006, 22:23-25 (emphasis added).

165. The updated solution vectors are passed to a “recombination operation section 108 by setting all the solution vectors in the solution vector set as the ***target*** of the recombination operation,” (EX1006, 22:67—23:11) (emphasis added), thereby updating the estimate of the target hyperparameter vector by using the random sample.

166. More specifically, Watanabe uses “a genetic algorithm processing section for performing the ***estimation*** of an optimal solution by genetic recombination operations;” which outputs “as the optimal solution vector a solution vector having the highest fitness in the most recent solution vector set.” EX1006, 22:21-31.

167. The recombination process uses a barycentric estimation method “for selecting a solution vector group P' having a higher fitness than a predetermined

reference fitness f^{th} , and for judging whether an **update direction vector along which vectors having high fitness are distributed can be *estimated*** from the barycentric vector g_1 (1 is the number of iterations) of the selected vector group P' and from a previous barycentric vector;”... and “**updating a plurality of solution vectors** along the direction of the update direction vector when the update direction vector is estimated....” EX1006, 22:39-57 (emphasis added).

168. Watanabe’s “update direction acquisition section” evaluates “the degree of matching between v_1 and v'_{1-1} stored in the update direction candidate recording section 110, and [], when it is judged that they match,... then passing the processing to the recombination operation section 108 by setting all the solution vectors in the solution vector set as the **target** of the recombination operation.” EX1006, 22:67-23:11 (emphasis added). This is updating the estimate of the target solution vector [i.e., hyperparameter vector].

169. The Examiner, in the Reasons for Allowance, states that “a random sample of the vectors is from a set of possible (i.e., candidate or hypothetical and not measured) hyperparameter vectors.” EX1002, p.24.

170. While I am not advocating a proposed claim construction based on the Examiner’s comments, the prior art nonetheless discloses the claimed subject matter even under this construction by the Examiner, as a candidate or hypothetical vector

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

is any that might result from Watanabe's recombination operation(s) as opposed to vectors that actually resulted from the recombination operation(s).

171. For instance, Watanabe discusses that its update direction candidate recording section obtains a difference vector, and records it as an update direction candidate vector. *See* EX1006, 24:58-25:24.

172. A random number technique (e.g., "Roulette wheel selection") is used for extracting solution vectors from a candidate solution vector set. *See* EX1006, 23:18-37; 26:17—27:8.

173. A barycenter shifting technique is used to update the vector by arbitrarily extracting points from a candidate of points contained in a sphere of multi-dimensional space, whereby neighboring vectors are randomly selected from among the solution vectors contained in the sphere. *See* EX1006, 25:59-65.

174. Thus, Watanabe's technique lays out how its random sample of vectors can be from a set of possible (i.e., "candidate or hypothetical and not measured") hyperparameter vectors.

175. The Examiner in the Reasons for Allowance states that "the claimed estimate variable is updated purely by using a random sample of vectors." EX1002, p.24.

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

176. While I am not advocating a proposed claim construction based on the Examiner's comments, the prior art nonetheless discloses the claimed subject matter even under this construction by the Examiner.

177. Watanabe discloses and suggests the sequence of claimed steps.

178. To the extent Claim 1 requires performing the claimed steps in a particular sequence, Watanabe discloses and suggests the sequence of steps contemplated in Claim 1. *Claim 1 encompasses a sequence of steps in which Element [1.3] is performed prior to Element [1.2].* This is because Element [1.4] explicitly recites that “the step of updating the estimate of the target hyperparameter vector” recited in Element [1.2] uses “said hyperparameter vector producing the best result” recited in Element [1.3]. That is, Claim 1 calls for selecting the hyperparameter vector producing the best result according to Element [1.3], and the selected hyperparameter vector is then used to update the estimate according to Element [1.2].

179. In addition, the '120 Patent describes its “method of determining hyperparameters [by]... each iteration comprising the steps of: drawing a random sample of hyperparameter vectors from a set of possible hyperparameter vectors, and updating the estimate of the target hyperparameter vector by using the random sample, the method being characterised in that **each iteration comprises** the step of

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

selecting from the random sample a hyperparameter vector producing the best result in the present and any previous iterations, and in that the step of updating the estimate of the target hyperparameter vector involves said selected hyperparameter vector.” EX1001, 2:1-20 (emphasis added).

180. As explained herein concerning Elements [1.2]-[1.3], Watanabe discloses and suggests performing the step of Element [1.3] prior to the step of Element [1.2].

**c. [1.3]: “selecting, from the random sample of
hyperparameter vectors, a hyperparameter vector
producing a best result in the present and any
previous iterations, and”**

181. Watanabe discloses and suggests Element [1.3].

182. As explained in Element [1.2], Watanabe discloses and suggests that its solution vector set P constitutes a “random sample of hyperparameter vectors” and “using the random sample” for “updating the estimate of the target hyperparameter vector” as claimed. *See* Section XIII.A.1.b.

183. It is also explained above that Watanabe uses a genetic algorithm to perform a recombination operation, which is updating the estimate of the target hyperparameter vector. *See* Section XIII.A.1.b.

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

184. Genetic-based classifier tuning algorithms utilizing crossover, mutation, recombination, and elitism techniques, had been used at the time of the '120 Patent for hyperparameter optimization. *See* Section XII.A.

185. It was well-known for recombination operations of genetic-based algorithms to include elitism selection (i.e., selecting a solution producing a best result in the present and any previous iterations), as evidenced by Shaefer (EX1013) (“The [genetic] operators may also include (i) an elitist operator which stores information about the token representing a favorable possible solution in one iteration, and reimports the information to a population of tokens in a later iteration; (ii) another elitist operator that adjusts the adaptive translation mapping to assure that information is not lost from one adaptation procedure to the next;...” 7:11-26.

186. An elitist strategy (e.g., the one disclosed by Shaefer), ensures the best of any previous iteration is in the next iteration.

187. Even if Watanabe’s recombination operation is somehow found to not expressly satisfy the recited “selecting, from the random sample of hyperparameter vectors, a hyperparameter vector producing a best result in the present and any previous iterations” (which it does), it would have been obvious to utilize an elitist strategy because elitism (which is precisely “producing a best result in the present and any previous iterations”) was a well-known technique of genetic algorithms.

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

188. A POSITA would have been motivated to use an elitist strategy for efficiently progressing through an update process to find the best solution vector.

189. Moreover, Shaefer demonstrates that elitism was a conventional technique employed in genetic algorithms, and therefore a POSITA would have been able to implement an elitism strategy using Watanabe's genetic algorithm in a predictable way with a predictable result.

190. In the Reasons for Allowance, the Examiner stated that the "best result" "is not produced after a later calculation or later use. It is the 'best result' instantaneously in the present". EX1002, pp.24-25.

191. Watanabe's solution vector having the highest fitness in the most recent solution vector set is the best result instantaneously in the present at least because it constitutes the best result with respect to the present and previous iterations, as explained above.

192. The Examiner also stated that "the vector produces a best result and is not accepted, itself, as the best result". EX1002, p.23.

193. Watanabe's "solution vector having the highest fitness in the most recent solution vector set" constitutes the best result with respect to the present and previous iterations, but also Watanabe employs separate criteria for termination. *See* EX1006, 27:31-46 (discussing termination conditions).

d. [1.4]: “wherein the step of updating the estimate of the target hyperparameter vector uses said hyperparameter vector producing the best result.”

194. Watanabe discloses and suggests Element [1.4].

195. Watanabe discloses “updating the estimate of the target hyperparameter vector by using the random sample”, “wherein the step of updating the estimate of the target hyperparameter vector uses said hyperparameter vector producing the best result”. *See* Section XIII.A.1.b/c. (discussing Elements [1.2]-[1.3]).

196. Element [1.4] recites that “the step of updating the estimate of the target hyperparameter vector” recited in Element [1.2] uses “said hyperparameter vector producing the best result” recited in Element [1.3].

197. That is, the hyperparameter vector producing the best result is selected according to Element [1.3], and the selected hyperparameter vector is then used to update the estimate according to Element [1.2].

198. Thus, Claim 1 at least encompasses performing the step recited in Element [1.3] before the step recited in Element [1.2], and Watanabe discloses and suggests performing the step recited in Element [1.3] before the step recited in Element [1.2]. *See* Section XIII.A.1.b/c.

2. **Claim 2: “The method according to claim 1, further comprising the steps of: selecting a further hyperparameter vector producing the best result in any previous iterations, and restricting the random sample of hyperparameter vectors by using the further selected hyperparameter vector.”**

199. Watanabe discloses and suggests Claim 2.

200. Watanabe discloses and suggests a step of selecting a further hyperparameter vector (*e.g.*, “past solution vector updates”) producing the best result in any previous iterations. EX1006, 27:52-55 (“At the same time, solution vectors are optimized by performing a recombination operation on the initial solution vector set, thus achieving fast and efficient optimal solution estimation *using the history of past solution vector updates*”).

201. Watanabe discloses and suggests a step of restricting the random sample of hyperparameter vectors (*e.g.*, “the solution vector group is updated” (*id.*, 27:49-50) by using the further selected hyperparameter vector (*e.g.*, “using the history of past solution vector updates”) (*id.*, 27:55).

202. As explained above concerning Claim 1, the Watanabe iterative process produces the best solution vector (*e.g.*, elitism via recombination of highest fitness solution vectors) at each iteration which is used to update the solution vector set. Watanabe specifically discusses an iterative process by which a solution vector set

is updated based on an elitism strategy (EX1006, 27:31-55), and thus discloses and suggests the steps of Claim 2.

3. Claim 3: “The method according to claim 2, wherein the step of restricting the random sample of hyperparameter vectors involves using an interval surrounding the further selected hyperparameter vector.”

203. Watanabe discloses and suggests Claim 3.

204. Watanabe discloses and suggests that the step of restricting the random sample of hyperparameter vectors (recited in Claim 2) involves using an interval surrounding the further selected hyperparameter vector.

First, the processing will be described when the update direction vector v has been obtained. In this case, **the process proceeds to the processing in the direction application updating section 107 [i.e., the further selected hyperparameter]**, where the processing is performed first in the barycenter shifting section 112 and then in the barycenter-surrounding solution vector generating section 113. The barycenter shifting section 112 performs the processing to shift the barycentric vector g_1 , calculated by the barycenter estimating section 109, in the direction in which the update direction vector v is pointing.... The barycenter-surrounding vector generating section 113 **arbitrarily extracts a number, a_{-s} , of points contained in a sphere of multi-dimensional space of radius a_{-len} with its center at the new barycentric vector g_{1+1} calculated by the barycenter shifting section 112 [i.e., interval surrounding the further selected**

hyperparameter vector], as shown by Expression 5, and sets these points as the members of a new solution vector set [**i.e., restricting the random sample of hyperparameter vectors**].

EX1006, 25:43-65 (emphasis added).

205. Watanabe's selects neighboring vectors from among solution vectors containing in a sphere, wherein Watanabe's recombination operation only operates on solution vectors within each of the neighbor vector groups, and therefore Watanabe discloses and suggests the step of "restricting the random sample of hyperparameter vectors involves using an interval surrounding the further selected hyperparameter vector" required by claim 3.

4. Claim 4: "The method according to claim 3, wherein the step of restricting the random sample of hyperparameter vectors is carried out prior to the step of selecting the hyperparameter vector producing the best result in the present and any previous iterations."

206. Watanabe discloses and suggests Claim 4.

207. Watanabe discloses and suggests that the step of restricting the random sample of hyperparameter vectors (recited in Claims 2 and 3) is carried out prior to (e.g., in a previous "iteration" of Watanabe's process) the step of selecting the hyperparameter vector producing the best result in the present and any previous iterations.

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

208. As with Claim 2, Watanabe's optimization process involves "estimation of an optimal solution vector [being] performed by *iteratively* performing the above process until either of the termination conditions, 1 or 2, is satisfied," whereby "the solution vector group is updated by estimating the direction in which solution vectors of high fitness are expected to be distributed [*i.e., restricting the random sample of hyperparameter vectors*]." EX1006, 27:31-55 (emphasis added).

209. This is performed *in each iteration*, and therefore is "*carried out prior to the step of selecting the hyperparameter vector producing the best result in the present and any previous iterations*" as claimed. EX1001, Claim 4.

5. Claim 5: "The method according to claim 3, wherein the step of restricting the random sample of hyperparameter vectors is carried out during the step of updating the estimate of the target hyperparameter vector."

210. Watanabe discloses and suggests Claim 5.

211. Watanabe discloses and suggests that the step of the restricting the random sample of hyperparameter vectors (recited in Claims 2 and 3) can also be carried out during the step of updating the estimate of the target hyperparameter vector (*e.g.*, "[a]t the same time, solution vectors are optimized by performing a recombination operation on the initial solution vector set" EX1006, 27:44-55)

212. Watanabe's updating procedure is described as follows:

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

“The estimation of an optimal solution vector is performed by iteratively performing the above process until either of the termination conditions, 1 or 2, is satisfied. In the present embodiment, as described, attention is focused on the shift vector of the barycentric vector of a group having high fitness within the solution vector set, and **the solution vector group is updated [i.e., restricting the random sample]** by estimating the direction in which solution vectors of high fitness are expected to be distributed. **At the same time, solution vectors are optimized by performing a recombination operation on the initial solution vector set**, thus achieving fast and efficient optimal solution estimation using the history of past solution vector updates.

EX1006, 27:44-55 (emphasis added).”

6. **Claim 6: “The method according to claim 2, wherein the step of restricting the random sample of hyperparameter vectors is carried out prior to the step of selecting the hyperparameter vector producing the best result in the present and any previous iterations.”**

213. Watanabe discloses and suggests Claim 6.

214. Claim 6 recites the same language as Claim 4 but Claim 6 depends from Claim 2, whereas Claim 4 depends from Claim 3.

215. Claim 4 requires the same process steps as Claim 6 but for the step of “restricting the random sample of hyperparameter vectors involves using an interval surrounding the further selected hyperparameter vector” which is required by Claim 3.

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

216. Watanabe's optimization process of "estimation of an optimal solution vector [which is] performed by iteratively performing the above process until either of the termination conditions, 1 or 2, is satisfied," whereby "the solution vector group is updated by estimating the direction in which solution vectors of high fitness are expected to be distributed" (EX1006, 27:31-55) satisfied the limitation required by Claim 6. *See* Claim 4, *supra*.

7. Claim 7: "The method according to claim 2, wherein the step of restricting the random sample of hyperparameter vectors is carried out during the step of updating the estimate of the target hyperparameter vector."

217. Watanabe discloses and suggests Claim 7.

218. Claim 7 recites the same language as Claim 5 but Claim 7 depends from Claim 2, whereas Claim 5 depends from Claim 3.

219. Therefore, Claim 7 requires the same process steps as Claim 5 but for the step of "restricting the random sample of hyperparameter vectors involves using an interval surrounding the further selected hyperparameter vector" which is required by Claim 3.

220. Therefore, Watanabe's updating procedure involving "the solution vector group is updated" and "[a]t the same time, solution vectors are optimized...." (EX1006, 27:44-55) satisfied the limitation required by Claim 7. *See* Claim 5, *supra*.

8. Claim 8: “The method according to claim 1, wherein the step of updating the estimate of the hyperparameter vector uses a weighting function.”

221. Watanabe discloses and suggests Claim 8.

222. Watanabe discloses and suggests updating the estimate of the solution vector [*i.e.*, hyperparameter vector] using a fitness function [*i.e.*, weighting function].

223. Watanabe’s use of its fitness function for updating its estimate of a solution vector is described as follows:

“Starting with this initial solution vector set P, the **estimation of an optimal solution vector** is performed. At the same time, the reference fitness f' to be used in the barycenter estimating section 109 and the reference shifting distance “length” to be used in the barycenter shifting section 112 are set. The evaluation value acquisition section 104, as previously described, calculates an evaluation value E for each solution vector p (k=1,..., n) using a **fitness function**, for example, as

$$E_k = \frac{(w(q_k) - w_{\min}^f)}{\sum_{k=1}^n (w(q_k) - w_{\min}^f)}$$

shown by Expression 2.

EX1006, 24:21-35 (emphasis added).

224. Watanabe’s estimation of an optimal solution vector is performed using a fitness function as a weighting function, and therefore Watanabe discloses and suggests “*updating the estimate of the hyperparameter vector uses a weighting function*” as required by Claim 8.

9. Claim 11: “The method according to claim 1, wherein the steps are carried out by a programmed computer apparatus including a processor and a computer readable medium including computer executable instructions.”

225. Watanabe discloses and suggests Claim 11.

226. Watanabe discloses and suggests Claim 11. Watanabe discloses and suggests the steps are carried out by a programmed computer apparatus (*e.g.*, “optimization adjusting apparatus using genetic algorithms” EX1006, 1:32-40) including a processor (“The main processing section 3203 comprises: a user evaluation section 3205 for presenting information expressed by each solution vector to the user and for making the user evaluate the solution vector;...” EX1006, 42:9-13; FIG. 32; see also 10:64-111:6; FIG. 41) and a computer readable medium (*e.g.*, “RECORDING MEDIUM, 3402” EX1006, 19:27) including computer executable instructions.

227. Watanabe identifies a “RECORDING MEDIUM, 3402” at col. 19, line 27. Watanabe also discusses use of its technique within a neural network – *see* Element [1.0], *supra*.

228. A POSITA would have understood that Watanabe discloses carrying out the steps of Claim 1 “by a programmed computer apparatus including a processor and a computer readable medium including computer executable instructions,” as is required by claim 11.

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

229. As evidenced by Friedrichs (EX1011), hyperparameter selection is commonly performed for classifiers in computers (e.g., computers having processor(s) computer readable medium/media). See EX1011, Abstract; 107:24-29; 112:13-25; 115.

230. A POSITA would have understood the recording medium,” main processing section,” and “evaluation section” disclosed by Watanabe to refer to the computer readable medium including computer executable instructions to be executed by a processor of a computer, which can be configured for performing classification via a classifier.

10. Claim 12: “A classifier for use in a machine learning system, the classifier using hyperparameters as control parameters, wherein the hyperparameters are determined according to the set of steps recited in claim 1.”

231. Watanabe discloses and suggests Claim 12.

232. Watanabe discloses and suggests a “classifier (e.g., Claim 11, *supra*) for use in a machine learning system (e.g., Genetic Algorithms in Search, Optimization, and Machine Learning” EX1006, 1:32-40).

233. Watanabe discloses and suggests the classifier using hyperparameters as control parameters through discussion of its technique within a neural network. See Section XII.A., Section XIII.A.1.

11. **Claim 13 (Preamble): “A non-transitory computer readable medium product including computer-executable instructions for carrying out a method of determining hyperparameters of a classifier in a machine learning system by iteratively producing an estimate of a target hyperparameter vector, each iteration comprising the steps of:”**

234. Watanabe/Friedrichs discloses and suggests the Preamble. *See* Claim 1 (Preamble), *supra*.

235. Regarding the non-transitory medium, classifier, and machine learning system *see* Element [11.0], *supra*.

- a. **[13.1]: “drawing a random sample of hyperparameter vectors from a set of possible hyperparameter vectors,”**

236. Watanabe/Friedrichs discloses and suggests Element [13.1]. *See* Element [1.1], *supra*.

- b. **[13.2]: “updating the estimate of the target hyperparameter vector by using the random sample, and”**

237. Watanabe/Friedrichs discloses and suggests Element [13.2]. *See* Element [1.2], *supra*.

- c. **[13.3]: “selecting, from the random sample of hyperparameter vectors, a hyperparameter vector producing a best result in the present and any previous iterations, and”**

238. Watanabe/Friedrichs discloses and suggests Element [13.3]. *See* Element [1.3], *supra*.

- d. **[13.4]: “wherein the step of updating the estimate of the target hyperparameter vector uses said hyperparameter vector producing the best result.”**

239. Watanabe/Friedrichs discloses and suggests Element [13.4]. *See* Element [1.4], *supra*.

- 12. **Claim 14 (Preamble): “A device for determining hyperparameters of a classifier in a machine learning system, the device comprising a processor arranged for iteratively producing an estimate of a target hyperparameter vector, each iteration comprising the steps of:”**

240. Watanabe/Friedrichs discloses and suggests the Preamble. *See* Element [1.0], *supra*.

241. Regarding a device comprising a processor, classifier, and machine learning system, *see* Elements [11.0]-[13.0].

- a. **[14.1]: “drawing a random sample of hyperparameter vectors from a set of possible hyperparameter vectors,”**

242. Watanabe/Friedrichs discloses and suggests Element [14.1]. *See* Element [1.1], *supra*.

b. [14.2]: “updating the estimate of the hyperparameter vector by using the random sample, and”

243. Watanabe/Friedrichs discloses and suggests Element [14.2]. *See* Element [1.2], *supra*.

c. [14.3]: “selecting, from the random sample of hyperparameter vectors, a hyperparameter vector producing a best result in the present and any previous iterations, and”

244. Watanabe/Friedrichs discloses and suggests Element [14.3]. *See* Element [1.3], *supra*..

d. [14.4]: “wherein the step of updating the estimate of the target hyperparameter vector uses said hyperparameter vector producing the best result.”

245. Watanabe/Friedrichs discloses and suggests Element [14.4]. *See* Element [1.4], *supra*.

13. Claim 15: “The device according to claim 14, wherein the processor is further arranged for: selecting a further hyperparameter vector producing the best result in any previous iterations, and restricting the random sample by using the hyperparameter vector producing the best result in any previous iteration.”

246. Watanabe discloses and suggests Claim 15. *See* Claim 2, *supra*.

14. Claim 16: “A machine learning system, comprising the device for determining hyperparameters of a classifier defined in claim 14.”

247. Watanabe discloses and suggests Claim 16. *See* Claim 14, *supra*.

B. Claim 9 Would Have Been Obvious Over Watanabe (EX1006) and Friedrichs (EX1011), in View of de Boer (EX1007) (Ground 2)

248. Claim 9 is obvious over Watanabe/Friedrichs in view of de Boer.

1. **Claim 9: “The method according to claim 1, wherein E^t is the selected hyperparameter vector X^t_i at iteration t producing the best result $S(X^t_i)$, the selected hyperparameter vector X^t_i having elements X^t_{ij} , wherein the step of updating the estimate of the target hyperparameter vector comprises the step of determining the hyperparameter v^t_j , where**

$$v^t_j = \frac{\sum_{i=1}^n I\{S(X^t_i) \geq \gamma^t\} W(X^t_i; E^t) X^t_{ij}}{\sum_{i=1}^n I\{S(X^t_i) \geq \gamma^t\} W(X^t_i; E^t)},$$

wherein γ^t is a threshold value and W is a weighting function.”

249. Watanabe/Friedrichs as applied to Claim 1, in view of de Boer, discloses and suggests Claim 9.

250. The formula of Claim 9 mathematically describes an iterative process as a “method similar to the cross entropy” for selecting a hyperparameter vector, each iteration producing the best result based on a threshold and a weighting function. *See, e.g.*, EX1001, 1:66; 2:27.

251. The '120 Patent states that “[t]he present invention allows a method **similar to the cross-entropy (CE) method** to be used for estimating and determining hyperparameters.” EX1001, 4:61-63 (emphasis added).

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

252. Watanabe likewise describes a similar application of iteratively processing solution vectors to find an optimized solution vector based on thresholds and weighting functions. Watanabe does not disclose the exact formula recited in Claim 9 for a neural computation. EX1006, 23:61 to 27:55.

253. de Boer's Algorithm 1.1 is an applied CE method for neural computation, and a notational variant of the CE parameter-update rule of Claim 9:

“In this case the ‘change of measure’ is determined by the parameter vector $v = (v_1, \dots, v_5)$. The main problem now is how to select a v which gives the most accurate estimate of ℓ for a given simulation effort. ... [A] quite general CE algorithm for rare event estimation is outlined below.

1. Define $\hat{v}_0 := u$. Set $t := 1$ (iteration counter).
2. Generate a random sample $X_1, \dots, X_N$ according to the pdf $f(\cdot; \hat{v}_{t-1})$. Calculate the performances $S(X_i)$ for all i , and order them from smallest to biggest, $S_{(1)} \leq \dots \leq S_{(N)}$. Let $\hat{\gamma}_t$ be the $(1 - \rho)$ sample quantile of performances: $\hat{\gamma}_t := S_{(\lceil (1-\rho)N \rceil)}$, provided this is less than γ . Otherwise, put $\hat{\gamma}_t := \gamma$.
3. Use the same sample to calculate, for $j = 1, \dots, n$ ($= 5$),

$$\hat{v}_{t,j} = \frac{\sum_{i=1}^N I_{\{S(X_i) \geq \hat{\gamma}_t\}} W(X_i; u, \hat{v}_{t-1}) X_{ij}}{\sum_{i=1}^N I_{\{S(X_i) \geq \hat{\gamma}_t\}} W(X_i; u, \hat{v}_{t-1})}. \quad ,,$$

EX1007, p.23; *see also* EX1007 in its entirety.

254. de Boer's CE method is suitable for neural computation (“An increasing number of applications is being found for the CE method. Recent publications on applications of the CE method include:... neural computation (Dubin, 2002...” EX1007, p.21.), and therefore a POSITA would have been

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

motivated to use de Boer's equation for Watanabe's neural computation of determining hyperparameters of a classifier.

255. A POSITA would have understood de Boer's CE formula as suited for updating the estimate of the target hyperparameter vector disclosed by Watanabe. Such updating renders the Claim 9 general formula for CE in neural computations obvious as de Boer describes its formula as a "general CE algorithm."

256. A POSITA would have been motivated to use de Boer's formula to optimize the updating of the hyperparameters in Watanabe, and would have incorporated de Boer's formula in Watanabe's system in a predictable way with a predictable result.

257. The '120 Patent attributes no criticality to the use of the claimed formula in its iterative process, or asserts any unique functionality beyond that which would have been provided by CE methods that were well known at the time of the invention. Rather, the '120 Patent concedes that cross-entropy methods such as that of Claim 9 were well known (EX1001, 1:34-59), incorporates by reference the equation disclosed by the de Boer paper (EX1007) in its own disclosure (EX1001, 1:39-42), and ascribes novelty not to the claimed equation itself, but rather to its use in connection with the purportedly novel iterative process—which, as explained above, is not novel at all in view of Watanabe.

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

258. Clearly, use of well-known, non-critical cross-entropy methods in connection with the '120 Patent's iterative process does not impart patentability.

259. Below is a table which establishes the notational correspondence of de Boer's equation with the variables of Claim 9:

120 Patent	de Boers
v_j^t	v_j
X_i^t	X_i
X_{ij}^t	X_{ij}
$S(X_i^t)$	$S(X_i)$
γ^t	γ

260. de Boer maps to Claim 9 as follows:

- a. t : iteration index.
- b. n : number of hyperparameter samples drawn per iteration.
- c. X_i^t : the i -th sampled hyperparameter vector at iteration t
- d. X_{ij}^t : the j -th component of the hyperparameter vector X^t
- e. v^t : the current estimate of the *target* hyperparameter vector at iteration t .
- f. v_j^t : the j -th component of v^t
- g. E^t : the *elitist* or best-performing hyperparameter vector found up to and including iteration t .

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

- h. $S(X_i^t)$: performance score (e.g. accuracy, negative loss) obtained when using hyperparameters X_i^t
- i. $I(\cdot)$: indicator function, equal to 1 when its argument is true and 0 otherwise.
- j. γ^t is threshold at iteration t ,
- k. $W(X_i^t; E^t)$: the weight of hyperparameter vector i in iteration t and the elitists hypervector, which are described further in Claim 10.

261. de Boer's equation estimates a target vector as the weighted average of the better performing vectors (as determined by S), where the weight is a function of the distance between these vectors and the elitist vector.

262. Claim 9 uses a superscript t to reference values in each iteration, and de Boer's algorithm iterates t time and applies the equation to different values in each of the t iterations.

263. The mathematical structure identically achieves a normalized weighted sum.

264. Both equations:

1. draw samples X_i ,
2. compute a score $S(X_i)$,
3. identify elite samples exceeding a threshold,

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

4. apply a weight $W(\cdot)$,

5. compute an updated parameter vector as a weighted average.

265. Although de Boer uses the best of the prior iteration while Claim 9 uses the best so far of any iteration, this is of no consequence.

266. The de Boer equation provides a broader means of homing in on a solution than that of Claim 9.

267. Claim 9's equation and de Boer's equation generate the same result if the best so far is the best in the current iteration when they are both within the same update region.

268. Watanabe states:

“To achieve the above objects, according to an eighth optimization adjusting method and optimization adjusting apparatus of the invention, the whole space is globally and roughly divided and solution vectors are updated, thereby roughly estimating a region which is likely to contain optimal solution vectors. Thereafter, **a stepwise convergence criterion by which to judge the convergence of solution vectors is gradually tightened and the update region is localized in a stepwise manner, to update solution vectors locally and thereby estimating an optimal solution vector.**”

EX1006, 7:3-12. (emphasis added).

269. Because Watanabe teaches elitism (see EX1006, 67:30-35) with a desire to localize the update region, A POSITA would have been motivated to modify the de Bore equation to match that of Claim 9.

270. Because a POSITA would have been motivated to use the de Boer formula in Watanabe's system, Claim 9 is unpatentable given that the mathematical structure of the Claim 9 formula is identical to that of de Boer's formula for achieving a normalized sum.

C. Claim 10 Would Have Been Obvious Over Watanabe (EX1006) and Friedrichs (EX1011), in View of de Boer (EX1007), Kirsten (EX1008), and Xu (EX1017) (Ground 3)

1. Claim 10: "The method according to claim 9, wherein the weighting function W is given by

$$w(X_i^t; E^t) = 1 - \frac{\sqrt{\sum_{j=1}^m (X_{ij}^t - E_j^t)^2}}{\sqrt{\sum_{j=1}^m (X_{ij}^t)^2} \sqrt{\sum_{j=1}^m (E_j^t)^2}}.$$

271. Watanabe/Friedrichs as applied to Claims 1 and 9, in view of de Boer, Kirsten, and Xu discloses and suggests Claim 10.

272. The '120 Patent references using a Euclidean distance function for the distance measure. de Boer similarly discloses using a Euclidean distance function.

273. Moreover, de Boer discloses a CE formula for applying multi-purpose cross-entropy techniques in neural computations. Discussed supra re: Claim 9.

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

274. It was well known for distance-based similarity measures to play key roles in machine learning when the '120 Patent was filed, and therefore a POSITA would have looked to de Boer for a similarity measure for determining a hyperparameter.

275. de Boer's method involves weighting based on distance. It would have been obvious to a POSITA to normalize Euclidean distance by dividing as a function of the magnitude of the vectors to arrive at the Claim 10 function W because the similarity weighted function places an emphasis on vectors closest to a target vector.

276. Kirsten discloses a weighting function for a first-order instance-based learner that a POSITA would have considered for use in the Watanabe/Friedrichs/de Boer combination.

277. Kirsten's nearest neighbor function is described as implemented with **any** elementary distance function. See EX1008, p. 219 ("some corresponding elementary distance function is called"); see also EX1008 in its entirety.

278. In Kirsten the similarity/weight is $\text{sim}(x,y) := 1 - \text{dist}(x,y)$. Kirsten's weighted average allocates an influence of farther-away neighbors using a formula of Claim 9 to compute a weighted average, wherein the weight depends upon the similarity:

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

$$h(x_{new}) := \frac{\sum_{e=(x,y) \in N} \text{sim}(x, x_{new})^a \cdot y}{\sum_{e=(x,y) \in N} \text{sim}(x, x_{new})^a}$$

EX1008, p.221.

279. The '120 Patent states: “when the Euclidean distance between E^t and X^t_{ij} is zero, the [Claim 10] weighting function W is equal to one.” EX1001, 3:26-28.

Values of Kirsten’s $\text{sim}(x, y) := 1 - \text{dist}(x, y)$ are in the range $[0,1]$ like the Claim 10 weighting function.

280. For instance, Kirsten’s formula $h(X_{new})$ is “a first-order instance-based learner applying the classical k-nearest neighbor (kNN) classifier in a relational representation, using the distance measure....Our distance measure takes on values in the range $[0,1]$, so we can equivalently work with the function: $\text{sim}(x, y) := 1 - \text{dist}(x, y)$ which ranges from 1 for identical objects to “0 with increasing distance between objects.” EX1008, p. 220. “Given a set of examples E and a new instance X_{new} from X to be classified... [t]he algorithm will then predict $h(X_{new})$ using the formula above. EX1008, p. 220.

281. Accordingly, a POSITA would have chosen Kirsten’s weighting function $h(X_{new})$ as the general CE formula of de Boer for placing higher weights to vectors closest to a target vector in a predictable way with a predictable result.

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

282. In so doing, a POSITA would have recognized that a Euclidean Distance calculated with a multiplication function in the denominator (as is prescribed by Claim 10) would have been obvious, as evidenced by Rui's (EX1017) Cosine Similarity multiplication function. This function would have been used as the elementary distance function called for by Kirsten's similarity formula.

283. A POSITA would have looked to modify the similarity function of Kirsten to emphasize weighting vectors closest to a target vector based on Rui. Indeed, the '120 Patent inventor described deriving Claim 10's equation using Euclidean distance (*See* Raaijmakers (EX1015), p.36.), wherein if D represents a distance function and S represents similarity, with $D = 1 - S$, then Euclidean distance, Cosine Similarity, and Manhattan distance equations define distance and similarity measures that were commonly chosen for classical machine learning at the time.

284. An operational constraint of the '120 Patent's weighting function is for "W [to be] a weighting function and $I\{S(X_i^t) \geq \gamma^t\}$ [to be] an indicator function which is equal to 1 if $S(X_i^t)$ greater than or equal to the threshold value γ^t , and else is equal to 0". EX1001, 3:11-15.

285. A POSITA would have understood Kirsten's similarity function as such a distance measure, as any normalized Euclidean Distance satisfies this constraint.

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

286. The '120 Patent states that “the step of updating the hyperparameters may involve a weighting function based upon a distance function, preferably a Euclidean distance function.” EX1001, 3:24-26.

287. Kirsten’s formula computes distances between first-order objects (i.e., not vectors) and is not a Euclidean distance formula.

288. Yet, Kirsten’s formula is more general than the formula in Claim 10 because it can compare numeric and discrete values as well as list and terms. When comparing two simple vectors of only numeric values, Kirsten discloses that “[f]or arguments of these basic types, some corresponding elementary distance function is called.” EX1008, p. 219.

289. Thus, Kirsten’s formula is like a normalized Manhattan Distance. *See* Horváth (EX1014).

290. Claim 10 specifies a weighting function W based on the Euclidean distance between a sampled vector X_i^t and the elitist vector E^t .

291. The numerator in Claim 10 is the Euclidean distance.

292. This Euclidean distance is discussed Xu (EX1017).

293. The denominator restricts the weight to be less than 1 by dividing by a function of the magnitude of the two vectors producing the relative (vs. absolute) distance.

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

294. Xu's Cosine Similarity function (see Table 1, p. 648) uses the exact same denominator for making the distance independent of vector length. EX1017, Table 1, p. 648; *see also* Xu in its entirety.

295. The '120 Patent attributes no criticality to the use of the claimed weighting function W in its iterative process.

296. Nor does the '120 Patent assert that the claimed weighting function W provides any unique functionality beyond that which would have been provided by other conventional functions well known at the time of the invention as evidenced by Kirsten.

297. As noted above, the '120 Patent concedes that cross-entropy methods were well known (EX1001, 1:34-59), incorporates by reference the de Boer paper (EX1007) as part of its own disclosure (EX1001, 1:39-42), and ascribes novelty not to use of a cross-entropy method itself, but rather to its use in connection with the purportedly novel iterative process—which, as explained above, is not novel at all in view of Watanabe.

298. Clearly, use of well-known, non-critical cross-entropy methods (and associated weighting functions) in connection with the '120 Patent's iterative process does not impart patentability.

XIV. SECONDARY CONSIDERATIONS

299. As discussed herein, all elements of the Challenged Claims were known in the art, and any differences would have been obvious to a POSITA based on the disclosures of the applied references and the knowledge in the art.

300. I am not aware of any secondary considerations that would alter my opinion.

301. Any purported evidence of secondary considerations that Patent Owner may present in this proceeding would be insufficient to overcome the strong evidence of obviousness of the Challenged Claims.

Declaration of Michael J. Pazzani In Support of
Petition for *Inter Partes* Review of U.S. Patent No. 8,438,120

XV. CONCLUSION

302. For the reasons set forth above, it is my opinion that all elements of the Challenged Claims of the '120 Patent are disclosed or suggested by the prior art. In my opinion, the Challenged Claims of the '120 Patent would have been obvious over the prior art.

303. In signing this Declaration, I understand that the Declaration will be filed as evidence in a contested case before the Patent Trial and Appeal Board of the United States Patent and Trademark Office. I acknowledge that I may be subject to cross-examination in the case and that cross-examination will take place within the United States. If cross-examination is required of me, I will appear for cross-examination within the United States during the time allotted for cross-examination.

304. I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code.

Date: March 31, 2026



Michael J. Pazzani