ISSN : 2165-4069(Online)
ISSN : 2165-4050(Print)

IJARAI

International Journal of Advanced Research in Artificial Intelligence

Volume 3 Issue 7

www.ijarai.thesai.org

A Publication of
The Science and Information Organization

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH IN ARTIFICIAL INTELLIGENCE

THE SCIENCE AND INFORMATION ORGANIZATION

www.thesai.org | info@thesai.org

Editorial Preface

From the Desk of Managing Editor…

"The question of whether computers can think is like the question of whether submarines can swim." — Edsger W. Dijkstra, the quote explains the power of Artificial Intelligence in computers with the changing landscape. The renaissance stimulated by the field of Artificial Intelligence is generating multiple formats and channels of creativity and innovation.

This journal is a special track on Artificial Intelligence by The Science and Information Organization and aims to be a leading forum for engineers, researchers and practitioners throughout the world.

The journal reports results achieved; proposals for new ways of looking at AI problems and include demonstrations of effectiveness. Papers describing existing technologies or algorithms integrating multiple systems are welcomed. IJARAI also invites papers on real life applications, which should describe the current scenarios, proposed solution, emphasize its novelty, and present an in-depth evaluation of the AI techniques being exploited. IJARAI focusses on quality and relevance in its publications.

In addition, IJARAI recognizes the importance of international influences on Artificial Intelligence and seeks international input in all aspects of the journal, including content, authorship of papers, readership, paper reviewers, and Editorial Board membership.

The success of authors and the journal is interdependent. While the Journal is in its initial phase, it is not only the Editor whose work is crucial to producing the journal. The editorial board members , the peer reviewers, scholars around the world who assess submissions, students, and institutions who generously give their expertise in factors small and large— their constant encouragement has helped a lot in the progress of the journal and shall help in future to earn credibility amongst all the reader members.

I add a personal thanks to the whole team that has catalysed so much, and I wish everyone who has been connected with the Journal the very best for the future.

Thank you for Sharing Wisdom!

Editor-in-Chief
IJARAI
Volume 3 Issue 7 July 2014
ISSN: 2165-4069(Online)
ISSN: 2165-4050(Print)

Editorial Board

Reviewer Board Members

CONTENTS

A More Intelligent Literature Search
- Shoulders of Giants

Michael G King, Alison Van Bree
Sunshine Health and Dental Superclinic
429 Ballarat Rd, Sunshine,
Victoria, Australia 3020.

***Abstract*—Although the topic of study relates to an environmental/health issue, it is the methodology described which serves to showcase an embryonic form of a new "more intelligent" protocol of search algorithm. Through the implementation of this algorithm, an extensive automated literature base yielded a single credible solution to a previously unsolved problem. Faced with a distressing but entirely unexplained incidence of birth defects, the proposed model of knowledge scavenging worked through acknowledged gaps in understanding of increased (phosphate) fertilizer, enabled the template of known facts regarding the interactions of phosphates with the processes of mammal (and other animal) growth, of metabolic function, and of neurological development, and delivered a causal model which would not, at least not easily, derive from current literature search methods. Illustrating the practical value of a step forwards in the design of intelligent literature search, the present study provides a candidate cause to explain a cluster of bovine deformity.**

Keywords—automated literature search; database; search algorithm; Craniosynostosis; fibroblast growth factor receptor

I. INTRODUCTION

The levels of animal production from these pastures are impressive. Yet we are not quite sure just why this happens. . . . . There are gaps in our understanding of soil/animal/plant relationships, particularly when soil fertility is increased and grazing pressure is intensified [1].

It is easy to acknowledge that there are "gaps in our understanding" of any, indeed of every, topic in the bio-related disciplines, however a more accurate description of "what is not presently known" might be that there is likely the totality of the required knowledge already discovered but located in a fragmented array of publications located in a variety of nominally discrete disciplines. Such an assertion acknowledges the entirely vast number of bio-medical-chemical publications being added to by millions of further publications each year (in the date range 2012 – 2013 the number of hits returned for PubMed data base is 2056910). Thus it would be arrogance, hubris, or at least out of step with Occam's razor to assume that whatever is being investigated actually includes "gaps" in total human, published knowledge. It is the more likely case that somewhere there is the missing information but it is not findable by conventional search. The present paper looks to this problem, introduces a strategy of solution, and provides a precise example illustrating the benefits of a more intelligent literature search.

Animal Birth Defects: Acorn Disease.

Recent examples of clusters of birth defects among animals (from 1% to 50% incidence in certain areas in Australia during the recent drought) had no obvious explanation. The defects listed include facial deformities and limb malformation. Genetic issues were investigated and eliminated. In short, the animal scientists remain baffled. Conventional investigative approaches might be summarized by the following steps:

a) investigating known information about animal defects, and

b) considering the information even more fully, in ever increasing detail.

This approach makes intuitive sense, however in the situation where there no likely candidate explanation of the problem, and when none seems to be emerging from the investigation, it makes sense to look outside the square. For the purposes of the present study "outside the square" might mean looking at human birth defects. It makes an interesting, even ironical, twist of laboratory processes to use investigations on humans to help solve animal disease. Even further "outside the conceptual square" would be an entirely different approach to the computer process of literature search – a methodology designed with the biochemical/medical sciences in mind.

Human Deformities studies

The study of animal birth defects differs from that of human deformities: the former is largely of technical interest, while the latter has an emotive overlay; the former is a bump in an established (or ignored, or even suppressed) statistical data stream, while the latter is something which is disturbingly abhorrent at any frequency. Given the affective incentive, research into causes and foci of action of candidate causative agents is likely to be more vigorous and extensive in the case of human birth defects. Therefore in considering the situation of a putative increase in bovine birth defects, it is appropriate to cast the knowledge net wider and trawl for answers in the human-based data.

The calves born under the currently-studied cluster of "acorn disease" features deficient limb development, as well as cranio-facial deformity. Corresponding to these anomalies deficient limb development in humans is included in studies of the broad range of conditions grouped under the heading of pre-natal rickets, while an understanding of the cranio-facial defects in the target calves may benefit from a review of Craniosynostosis in humans. This review will take the two above-named domains of study in reverse order.

Craniosynostosis Studies

There are two putative forms of Craniosynostosis, both forms being conventionally linked to genetic mutations:

The Saethre-Chotzen and the Muenke craniosynostoses,. . . . Patients with Saethre-Chotzen and Muenke syndromes carry mutations in the TWIST and fibroblast growth factor receptor (FGFR) 3 genes, respectively [2].

And pursuing the mutant Growth Factor (FGF, FGFR etc) variant of this disorder, it is noted that:

Mutations in the FGFR gene family have been linked to a series of syndromes (the craniosynostosis syndromes) whose primary phenotype involves aberrant development of the craniofacial skeleton. . . . Unfortunately, studies attempting to link expression of mutant FGFRs with changes in cellular phenotype have yielded conflicting results [3].

At first glance, human Craniosynostosis studies would appear to have little potential to cast light upon the current acorn-diseased calves: first the human studies are "conflicting" and secondly, genetic factors in the diseased calves have been all but ruled out by veterinary review (private communication from investigating group). However there are many physiological and biochemical steps between genetic information and limb or facial formation, and when the causal emphasis upon mutant genes is put to one side, there remains a credible common thread: whether genetic or not, perturbations are found in the Growth Factor component of the complex process of bone development. Observe that the following quotation reads equally well whether or not genetic mutation is implied: Craniosynostosis syndrome-linked FGFR mutations have been shown to be gain of function in terms of receptor activation and have been presumed to result in increased levels of FGF/FGFR signalling [3].

Although Hatch's [3] specific proposition is that "increased levels of FGF/FGFR signalling" is the cause of Craniosynostosis, it would be sensible to keep the net wider and work under the more general rubric of "Growth Factor(s)", naming neither "fibroblast Growth Factor" (FGF) nor the "Receptor" (FGFR) as the specific focus of attention.

In summary, to this point, it appears relevant to search the human literature for similar or parallel defects in the quest for an understanding of bovine birth defects (acorn disease). Furthermore, it follows from an open-minded reading of the Craniosynostosis literature that the search could credibly consider links to Growth Factor (including specific chemical effects thereon), and also other studies of normal or abnormal limb and cranio development.

In seeking this "chemical cause", it was the underlying a priori position, that the chance of a "rare" causative agent in the sense of truly unusual or outstandingly complex was all but eliminated by definition. If there were a rare agent at work, its very rareness would bring it to the forefront of likely candidate causes - its rareness would be its downfall. Given the apparently mysterious nature of clusters of craniosynostosis, accepting that an answer has never immediately come to hand then the likely candidate cause logically should be some agent which will be masquerading as normality. And then following this position, the implicit "normality" of the causal factor leads to the expectation that the topic is well covered in the literature, albeit without the knowledge or recognition of the authors (and hence the literature search systems).

II. METHOD

The entire biochemical/biomedical data bases contain information from disparate sub-disciplines (branches of medicine and branches of the biochemical sciences). This knowledge base which well exceeds 100,000 items is credibly regarded as the work of *Giants*. The bio-medical data base is approached for specific "problems". These problems require an "answer", where the notion of an "answer" is conceptually regarded as a chemical agent which is either a causal vector, or an intervention. For the Shoulders of Giants (SOG) project it is the assumption that the answer is mentioned somewhere in this total literature base, along with oblique cues which do not directly link this agent to the notional "problem".

The hypothesis of the SOG algorithm is that the target information (a "candidate answer") may, at any stage of a literature search, be only one step away. The problem for the investigator is that, by definition, the direction of that step is neither known nor knowable. If the direction were known, then the investigator would immediately guide a conventional search in that direction.

The "conventional" computer based search is set to trawl data bases for instances of a target word or phrase. Thus, on the bio-medical data bases the word SMOKING occurs among perhaps 25,000 abstracts. The search is refined, in the end, by pairing the target with another limiter (for example ASTHMA, itself having, say, 16,000 hits) thus SMOKING and ASTHMA occurs only in only 2688 abstracts – still too many to review. Further limitations will bring this number down to a manageable size (notionally around 100 abstracts) which is hopefully highly focused (example SMOKING, ASTHMA and CAFFEINE might feature together in a human-perusable set of abstracts rather less than 100 in number).

Modern search strategies assist the search for parallel studies, by locating matches with the key words of other papers in the data base. These assisted searches are not designed to find an answer to a problem when the "answer" is as-yet unknown (unknown in the sense of not been linked to the key words of the "question" underlying the seeding search).

The following observations derive from a word-by-word review of around 30,000 abstracts from the medical science. The process employed was a dedicated but flexible computer program designed precisely for application in the medical/biological sciences where "causal links" are sought from an array of knowledge too large for normal contemplation. In this instance the program seeks to locate common factors linked to the human defect Craniosynostosis this factor being shared with other birth defects in animals. A shared putative causal factor would tend to support the notion of environmental cause.

The (SOG) project has devised a search strategy, or "algorithm", which can suggest answers even though these offerings have not been previously linked to the question – at

least not linked by bridges which are readily discovered through normal search processes.

The SOG project comprises a series of computer programs designed to:

a) Interact with the entire bio-medical data bases (BDB) and extract multiple large sub-sets of literature abstracts

b) isolate these large subsets of the BDB (perhaps thousands of abstracts each) for closer scrutiny

c) take the results of (b) and compare two or more notionally disparate subsets of BDB for overlap

d) provide a manageable few candidate "answers" (really "directions" of likely fruitful investigation) which can be followed by conventional investigative processes.

e) *At the current phase, collect together and print in RTF format (from the "extracted" data sets - step "a", above) sets of abstracts which feature a target "term" (usually a single word or phrase). This step is, of course, available from conventional data search (for example "Smoking" papers which mention "Asthma")*

Step "c" – comparison of two data sets.

This step is at the heart of the SOG program suite. Ultimately pairing for meaningful matches is the core process to deal with the potentially explosive output derived from conducting multiple, nominally unrelated, literature searches.

Once two data sets of any size (eg 20,000 abstracts) have been nominated (at present by human-initiated keyboard command (but in the future by an algorithm itself generated from the initial "problem") then the program can remove a pre-determined list of "trivial" terms, and review the two filtered data sets for meaningful overlap.

III. Results

First common Term.Review of the first cut of the literature (around 30,000 abstracts), points to a simple dichotomy: either the answer remains unknown, or the causative vector contains the term "phosphate". When all "trivial" terms were eliminated from the abstracts reviewed, this term was the only common word which linked the abstracts which derived from the various domains.

The literature review tendered the term "phosphate" as the only common word linking the various abstract groups. Taking the positive view that this link is "non-trivial" and therefore worth considering, the probable cause is some level of organo-phosphates ingested (exposed?) pre-partum. The relevant literature could now be searched in a more conventional manner using standard abstract-sifting strategies (eg "phosphate" with "craniosynostosis") or the down-loaded gamut of abstracts could be further trawled using the above-mentioned abstract-reviewing program.

Use of First Common Term A new and more focused review of the literature confirmed the non-controversial direct effects of various organophosphates, and the more indirect effects of putative metabolites of organophosphates (including the established potentially deleterious effects of deviant levels of "normal" phosphate metabolites) are noted in the literature.

A future review could, by currently-available search methodologies, collect and display all relevant references relating to:

a) the overlap of animal defects and craniosynostosis pointing to "phosphate" as the common factor;

b) the established links between FGF and craniosynostosis, including the relatively weak genetic explanation for FGF-related forms of the disorder: Mutations in the FGFR gene family have been linked to a series of syndromes (the craniosynostosis syndromes) [3].

c) *the established links between organophosphates and altered FGF activity.*

IV. Discussion

The putative organophosphate-linked explanation of craniosynostosis raises questions about the delivery vector, and the identity of the putative phosphate. It is relevant that the application of extra phosphate has been lauded as appropriate during drought years, with the goal of enhancing otherwise impoverished crops. It is already noted in the pro-phosphate agri-literature that the precise mechanisms of this method are not clearly understood – and that "gap in understanding" referred only to the crop-producing qualities. Little or no consideration appeared to be accorded to effects of additional phosphates outside the domain of grass growth.

The current explanation (that increased phosphate is a causal factor in birth defects) has the advantage of matching at least one of the conceptual conditions precursive to the present study: it was anticipated that a genuinely novel chemical compound would NOT have been the cause of the defects, and more particularly, that a popular and presumed safe compound must be the culprit.

V. Conclusion

Noting that the final validity of the science under scrutiny is not the central issue, but merely an illustration of the potential results from a step forward in intelligent search algorithms, it is presently submitted that a credible *a priori* case for organophosphate activity as the causal factor of both craniosynostosis and animal birth defects exists. This proposition is based upon non-controversial and established literature however the proposed "answer" would not be arrived at by conventional literature search alone. The underlying search strategy firstly accepts that an entirely huge and unwieldy quantity of scientific investigation has been carried in an ever widening and ever more isolated myriad of separate fields.

It is the primary and founding proposition for the presently developed approach that almost any "advance" in understanding, even in the development of treatments and cures, will be seen (with the benefit of hindsight) to have touched on and brought together ideas which have been described in the literature – but these separate components have not been linked together. That is to say that many future (especially bioehcmical) developments will have been hiding in plain sight.

For the present first step (illustrating that a tentative solution can be found for a problem which has defied solution by

conventional focused research by experts), the elegance of the uncovered "solution" is firstly that a mundane cause for presumed genetic conditions has not previously been signaled nor sought, and (reversing the vectorial sense of the research) the studies of phosphate and limb or other development had not been interpreted as candidate explanations for putative genetic conditions. The reason why this answer has not been envisaged by others is because of the (likely) assumption that a mutagenic agent rather than an FGF-interactive agent was sought. The SOG algorithm makes no assumptions but merely seeks links where none were previously imagined; it is then up to the human scientist to consider the credibility and ultimately the validity of the suggested links.

In summary the SOG-derived propositions were:

- Cranial malformation at birth can be due to fibroblast (fibroblast growth factor) aberrations.
- Phosphates are involved in aberrant pre-natal development, possibly due to an interaction with fibroblast growth factor. Most common studies relate to low phosphate levels and bone deformities (hypophosphate conditions) however high serum phosphate levels are equally linked to relevant deleterious outcomes [4,5] of hyperphosphatemia and impaired skeletogenesis.
- Broader ranges of experiments beyond the neo-natal stage show a variety of interactions between phosphates (typically organophosphates) and various fibroblast or fibroblast growth factor activity.
- Prenatal exposure to aberrant phosphate conditions can lead to a fibroblast-mediated condition of rickets.

The strength of the present proposition lies in that it is based upon non-controversial science. In addition, there is not an obvious competing explanation for the coincidence of these disorders.

While it may be of interest to have a putative explanation of a mysterious cluster of animal or human birth defects, the true strength of the present study is that it showcases a novel method of investigation that was precisely designed to serve the bio-medical domain. The validity of the analytical procedure depends entirely upon the proposition that somewhere, in two or more nominally disparate domains, there already co-exist studies which can be brought together to cast light upon currently mysterious questions.

The illustrated "solution" (whether ultimately accepted or rejected) starkly shows a solution which would be invisible to current literature search algorithms which are not based upon the approach of comparing two or more disparate data sets in order to find the "hidden, but existing" solution to problems.

Future work would ideally build the "next step" into the automated search process. The process as applied to the present problem analysed the entire (that is all that were available as abstracts) literature base pertaining to a single search term, and then proceeded to list the frequency of any "non-trivial" words/phrases. This resulting list of over 4000 items was matched with another list of over 11,000 items and the resulting overlap of less than 300 items was compared by human consideration and the result was just one potential explanatory term. The next phase of the SOG project would be to have the computer program "learn" more about filtering non-explanatory terms. However the final strategy which is at this stage beyond the resources of the authors would be to enter a topic of interest (effectively a single search term) and the SOG suite automatically access the data bases, extract the likely immense set of literature, peruse these for terms worth pursuing, and follow with as many "matching search result sets" as necessary. The principal author would be pleased to cooperate in whatever manner may lead to the fruition of this new concept, while, returning to what can be learned from the present illustration of part of this whole strategy effectively done "by hand": in many cases, the truth is already out there, but in fragmented forms in diverse places.

REFERENCES

[1] Sale, P. *The Pasture Productivity Revolution*. Department of Agricultural Sciences. LaTrobe University, Melbourne, Australia. 12pp. 2007.

[2] Jadico SK. Huebner A. McDonald-McGinn DM. Zackai EH. Young TL. Ocular phenotype correlations in patients with TWIST versus FGFR3 genetic mutations. *Journal of AAPOS: American Association for Pediatric Ophthalmology & Strabismus.* 10(5):435-44, 2006 Oct.

[3] Hatch NE. Hudson M. Seto ML. Cunningham ML. Bothwell M. Intracellular retention, degradation, and signaling of glycosylation-deficient FGFR2 and craniosynostosis syndrome-associated FGFR2C278F. *Journal of Biological Chemistry*. 281(37):27292-305, 2006 Sep.

[4] Garringer HJ. Fisher C. Larsson TE. Davis SI. Koller DL. Cullen MJ. Draman MS. Conlon N. Jain A. Fedarko NS. Dasgupta B. White KE. The role of mutant UDP-N-acetyl-alpha-D-galactosamine-polypeptide N-acetylgalactosaminyltransferase 3 in regulating serum intact fibroblast growth factor 23 and matrix extracellular phosphoglycoprotein in heritable tumoral calcinosis. *Journal of Clinical Endocrinology & Metabolism*. 91(10):4037-42, 2006 Oct.

[5] Sitara D. Razzaque MS. Hesse M. Yoganathan S. Taguchi T. Erben RG. Juppner H. Lanske B. Homozygous ablation of fibroblast growth factor-23 results in hyperphosphatemia and impaired skeletogenesis, and reverses hypophosphatemia in Phex-deficient mice. *Matrix Biology*. 23(7):421-32, 2004 Nov.

Classifications of Motor Imagery Tasks in Brain Computer Interface Using Linear Discriminant Analysis

Roxana Aldea
Faculty of Electronics, Telecommunications and Information Technology
"Gheorghe Asachi" Technical University of Iasi
Iasi, Romania

Monica Fira
Institute of Computer Science
Romanian Academy
Iasi,Romania

***Abstract*—In this paper, we address a method for motor imagery feature extraction for brain computer interface (BCI). The wavelet coefficients were used to extract the features from the motor imagery EEG and the linear discriminant analysis was utilized to classify the pattern of left or right hand imagery movement and rest. The performance of the proposed method was evaluated using EEG data recorded by us, with 8 g.tec active electrodes by means of g.MOBIlab+ module. The maximum accuracy of classification is 91%.**

Keywords*— *Brain computer interface; motor imagery; wavelet; linear discriminant analysis

I. INTRODUCTION

Brain computer interface is a system of communication with the external environment, a device that reads brain signals and converts them into control and communication signals. The research on BCI domain is motivated by the hope of creating new communication channels for people with severe neuromuscular disabilities.

BCI can offer the patients who suffer from some diseases, like amyotrophic lateral sclerosis or total paralysis ("locked-in" syndrome), the possibility to communicate with the environment, to control computers, or to drive external devices by regulation produced by brain activity alone, [1].

In the 60's, the control of devices using brain signals was considered science fiction. Although recording brain signals have attracted attention since 1922, when the German scientist, Hans Berger [2], recorded the electrical activity of the brain, measurement technology and signal processing were still quite limited to understand how the brain operated. Nowadays the situation has changed. Research in the field of neuroscience in recent years has led to a much better understanding of the human brain. Algorithms and signal processing capabilities of computers have advanced so much that the real-time processing of signals from the brain not require expensive and very bulky equipment.

The movement of a member or even a single muscle contraction causes changes in brain activity. In fact, only the imagining or preparing of a movement modifies the sensorimotor rhythms.

Sensorimotor rhythms (SMR) refer to oscillations recorded on brain activity in somatosensory and motor areas. Brain oscillations are usually classified according to specific frequency bands, named: delta <4 Hz, theta: 4-7 Hz, alpha 8-12 Hz, beta: 12-30 Hz, gamma :> 30 Hz. Alpha rhythm activity recorded on sensorimotor areas is called the mu rhythm. The decrease in oscillatory activity in a specific frequency band is called event related desynchronization – ERD, [3]. Similarly, the increase of oscillatory activity in a specific frequency band is called event related synchronization - ERS. The patterns ERD / ERS can be produced by motor imagery. So, the sensorimotor rhythms are represented by mu (8-12 Hz) and beta rhythms (12-30Hz).

Imagining left hand movement produces a desynchronization on C4 electrode in the right side of the scalp, while imagining right hand movement produces a desynchronization on electrode C3, on the left side of the brain. The cerebral activity caused by hand movement is localized in the contralateral area of the brain.

In this paper we used multiresolution wavelet analysis for feature extraction. This method was very used in signal processing of BCI data recordings, [4], but in combination with linear discriminant analysis (LDA) we obtain a better classification rate than the classification obtained in the online cursor movement task.

The multiresolution wavelet analysis gives us a time localization of spectral components so time-frequency analysis represents a suited tool to get appropriate features which will be used to train the classifier.

The signals are classified using two methods based on LDA. We use this type of classifier because we want to compare our result with those obtained by the BCI2000 software, when it uses LDA to show the success of the testing paradigm. Our contribution is represented by the use of the LDA with the normalized feature matrix.

The goal of this paper is to show that the classifier used in BCI2000 can be improved to obtain better accuracy and our acquired signals are appropriate to be used to control a BCI system.

II. METHODOLOGY

A. Experimental paradigms

The EEG signals used for this experiment were recorded by means of a g.tec acquisition system, namely g.MOBIlab+ module, and BCI2000 platform. The data were recorded with 8 wet active electrodes, placed on scalp according to the international 10-20 system, [5].

The electrodes are placed on channels: CP3, CP4, P3, C3, Pz, C4, P4 and Cz. These channels are selected in both hemispheres, in sensorimotor areas, due to the appearance of sensorimotor rhythms in these areas. The reference electrode is placed on the right ear.

Train paradigm

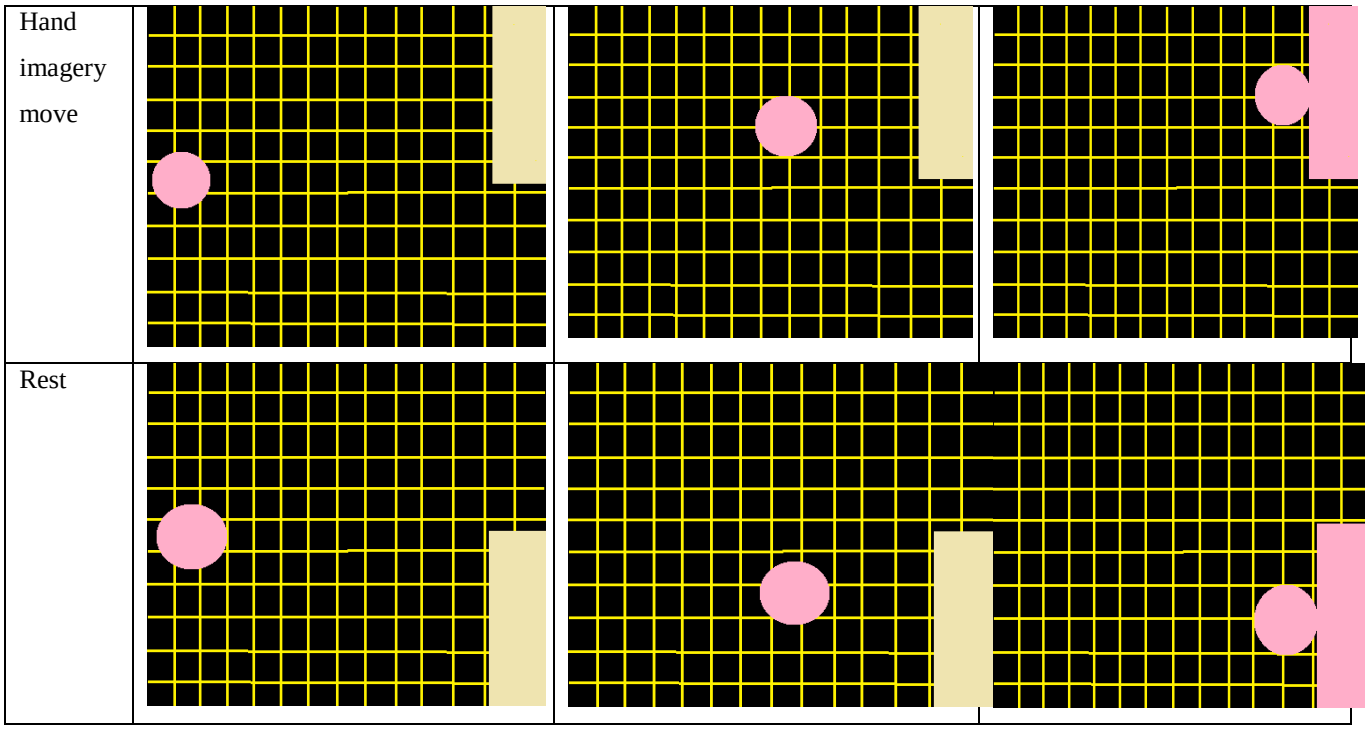
The subjects received instructions regarding their behavior during recording. The subjects were seated in front of a monitor that during the sessions will either be blank or displaying an arrow pointing left or right. When a left or right arrow is displayed, the subjects need to imagine the movement of the respective hand. When the screen monitor is blank, they must relax and stop any movement. Each left and right arrow appears 30 times. The time interval of the visual stimulus was 2 seconds. After this part of training we perform an offline analysis which computed the coefficient r^2 comparing the EEG spectra associated with each motor-imagery task with spectra recorded at rest, [6].

Test paradigm

During the testing paradigm the subject should imagine the movements of only one hand, that for which we obtained the best results in the offline analysis, such classification will be perform only for two classes: motor imagery of the left or right hand and rest.

On the testing paradigm the subject must lead a ball so that it hit the target, represented by a yellow bar. When the target is at the top of the screen, the subject must imagine the movement of the hand, and when the target is at the bottom of the monitor, the subject needs to relax. When the ball reaches the target, it changes color (Table I). At the end of the paradigm, BCI2000 software displays the percentage of success of the experiment (classification which is based on LDA).

TABLE I. TESTING PARADIGM



B. Feature extraction using multiresolutin wavelet analysis

Frequency analysis using Fourier transform represents a current method used to analyze EEG signals, because the spectral components of the SMR may contain useful information. Usually, some features of interest are found especially in the frequency bands within 0-60Hz domain. The Fourier transform highlights only the information concerning the spectral components revealed in the signal; it doesn't present the time localization. The localization in time of spectral components may be performed by means of time-frequency analysis such as Short Time Fourier transform (STFT) or the continuous and discrete wavelet transform.

In discrete time domain, digital filters with different cut off frequencies are used to analyze the signal at different scales. The signal is passed through a series of high pass filters to analyze high frequencies and through a series of low pass filters to analyze low frequencies.

The signal resolution (a measure of detail information carrier) changes by filtration and the scale by subsampling

(decimation). Subsampling by a factor, *n*, reduced the number of samples *n* times, [7].

The discrete signal, denoted by *x(n)* is passed thought a low pass filter, that cuts the superior half of the signal frequency band. The impulse response of the filter is *h(n)*. The filtration is equivalent to the signal convolution with the impulse response of the filter. In discrete-time, convolution is defined as, [8]:

$$y\ n = x\ n * h\ n = \sum_{k=-\infty}^{\infty} x\ k\ h\ n-k \qquad (1)$$

The multiresolution decomposition of a recorded signal is schematically shown in Figure 1.

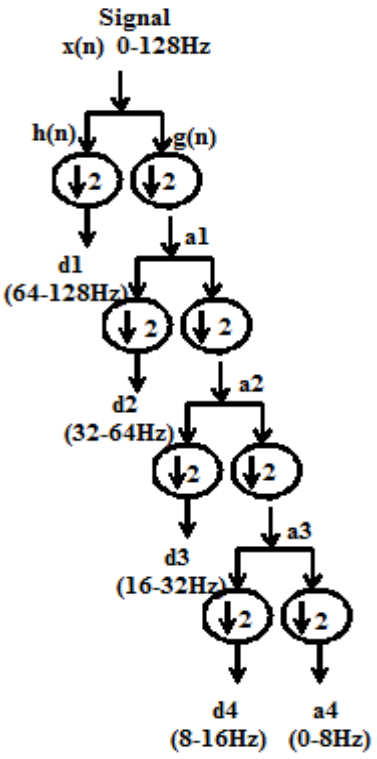


Fig. 1. Fourth level multiresolution wavelet decomposition

Taking into account that the frequency components of the EEG signal are in the 0-128Hz range, while the spectrum of the mu rhythm is around 8-12Hz and beta rhythm around 12-30Hz, a fourth level decomposition of the signal was required.

After the first level of decomposition, the EEG signal is decomposed in the detail coefficients of high frequency $\mathbf{D_1}$ (64-128Hz) and the approximation coefficients of low frequency $\mathbf{A_1}$ (0-64Hz). At the second level of the decomposition, the coefficients $\mathbf{A_1}$ are further decomposed in the detail coefficients $\mathbf{D_2}$ (32-64Hz) and approximation ones $\mathbf{A_2}$ (0-32Hz). Following this procedure, the coefficients $\mathbf{D_3}$ (16-32Hz), $\mathbf{A_3}$ (0-16Hz) and $\mathbf{D_4}$ (8-16Hz) and $\mathbf{A_4}$ (0-8Hz) are obtained.

The multiresolution decomposition is realized with Coiflet4 wavelet, [9], on C3, CP3, P3, C4, CP4 and P4 channels.

For linear discriminant analysis classification we use only the feature from the coefficients of interest: the detailed coefficient of fourth level with 8-16 Hz frequency band (corresponding to mu rhythm) and the detailed coefficient of third level decomposition with 16-32 Hz frequency band (corresponding to beta rhythm).

C. Linear discriminat analysis (LDA)

We used LDA classifier because it is one of the most effective linear classification methods for BCI and because it is also used by BCI2000 software on the testing paradigm. The method we used is a bit different applied and we want to compare the results with those obtained after the online paradigm.

LDA computes the discriminant vector $w \in$ that separates the classes best possible. Suppose we have a set of m samples $x_1, x_2, .., x_m$ belonging to a class. The objective function LDA is as follows, [10]:

$$a^* = \arg_a \max \frac{a^T S_b a}{a^T S_w a}, \qquad (2)$$

$$S_b = \sum_{k=1}^{c} m_k \ \mu^{(k)} - \mu \ \ \mu^{(k)} - \mu^{\ T}, \qquad (3)$$

$$S_w = \sum_{k=1}^{c} \left(\sum_{i=1}^{mk} x_i^{(k)} - \mu^{(k)} \ \ x_i^{(k)} - \mu^{(k)\ T} \right), \qquad (4)$$

where μ is the total samples vector, *mk* is the number of samples in the k-th class, $\mu^{(k)}$ is the average vector of the k-th class, and $x_i^{(k)}$ is the i-th sample in the k-th class. We call S_w the within-class scatter matrix and S_b the between-class scatter matrix.

Define $S_t = \sum_{i=1}^{m} (x_i - \mu)(x_i - \mu)^T$ as the total scatter matrix and we have $S_t = S_b + S_w$. The objective function of LDA in (2) is equivalent to:

$$a^* = \arg_a \max \frac{a^T S_b a}{a^T S_t a}, \qquad (5)$$

The optimal *a*'s are the eigenvector corresponding to the non-zero eigenvalue of the generalized eigen-problem:

$$S_b a = \lambda S_t a, \qquad (6)$$

Since the rank of S_b is bounded by c-1, there are at most c-1 eigenvectors corresponding to non-zero eigenvalues.

The basic idea of LDA is simple: a linear function of attributes is computed for each identified class. The class function with the highest score represents the predicted class.

There are many linear classification models and they differ greatly on how the coefficients are set. A quality of LDA is that it does not require multiple passes over the data to obtain optimization. LDA also faces up to problems with more than two classes, obtaining probability estimates for each of the classes.

III. RESULTS

The classification will be performed on those three channels corresponding to imagining movement of the right hand, C3, CP3 and P3, respectively C4, CP4 and P4 for motor imagery of the left hand.

The components of the features matrix were selected from the detailed coefficient of fourth level with 8-16 Hz frequency band and the detailed coefficient of third level decomposition with 16-32 Hz frequency band. This features matrix is computed for the training set and for the test set of the signals. Classification is performed between two classes: the relaxation and the imagined movement.

BCI2000 software uses a LDA classification method. The accuracy percentage is displayed at the end of the testing paradigm. We can observe than the paradigm is not so rigorous because the subject correctly imagine the movement or the relaxation but the ball did not reach the target. Because of this, we implemented our own classification software, in MATLAB and we obtained better classification accuracy.

We perform the classification with the LDA software implemented by us, for all the subjects and the result are expressed in percentage of accuracy. Then we classified the signals with a LDA MATLAB code that uses the normalization of the features matrix.

In TABLE II and III are presented the results obtained with the three classification methods when subjects imagine the right hand movement (Table I) and left hand respectively (Table II).

In TABLE II we obtain better classification rates with the LDA implemented than the LDA used by BCI2000 software, except one subject for which we obtain the same percentage. Also, we can see that for most subjects we obtained higher classification rates with the normalized features matrix LDA. The best classification rate is obtained, 86%, with the normalized LDA on channel P3.

TABLE II. LDA CLASSIFICATION CORRESPONDING TO THE MOTOR IMAGERY OF THE RIGHT HAND SIGNALS VERSUS THE REST SIGNALS

Sub.	LDA BCI 2000	C3		CP3		P3	
		LDA	*LDA norm*	*LDA*	*LDA norm*	*LDA*	*LDA norm*
1	45%	55%	**73%**	73%	**77%**	77%	**82%**
2	77%	68%	**82%**	82%	82%	77%	**82%**
3	72%	**82%**	77%	**82%**	77%	**77%**	68%
4	63%	77 %	77%	77%	**82%**	**82%**	77%
5	72%	72%	72%	73%	73%	73%	73%
6	54%	77%	**82%**	73%	77%	77 %	**82%**
7	77%	77%	77%	82%	82%	82%	**86%**
8	77%	**77%**	72%	68%	**77%**	77%	77%
9	77%	77%	**82%**	68%	**77%**	77%	73%

In TABLE II we obtain better classification rates with the LDA implemented than the LDA used by BCI2000 software for all the subjects. The best classification rate, 91%, was obtained with LDA classifier implemented by us, for the first subject on channel C4 and with LDA with the normalized features matrix on channel P4. We have achieved a better classification, for most subjects, when we used LDA with the normalized features matrix except subject 15 on channel C4 and subject 11 on channel P4 when we obtain better classification with LDA.

From the results in both tables we can observe that we got better results when we use LDA classification methods that we implemented, compared to those obtained with the BCI2000 software. From the two LDA methods used, we have achieved a better classification, for most subjects, when we used LDA with the normalized features matrix.

TABLE III. LDA CLASSIFICATION CORRESPONDING TO THE MOTOR IMAGERY OF THE LEFT HAND SIGNALS VERSUS THE REST SIGNALS

Sub.	LDA BCI 2000	C4		CP4		P4	
		LDA	*LDA norm*	*LDA*	*LDA norm*	*LDA*	*LDA norm*
10	63%	**91%**	86%	86%	86%	86%	**91%**
11	86%	73%	**82%**	77%	**82%**	**82%**	77%
12	63%	68%	**82%**	73%	**77%**	68%	**73%**
13	45%	68%	**82%**	68%	**77%**	77%	**82%**
14	60%	77%	77%	77%	**82%**	77%	77%
15	68%	**86%**	77%	77%	77%	77%	77%
16	77%	77%	77%	73%	73%	77%	**82%**
17	77%	77%	**82%**	73%	**82%**	77%	73%

IV. CONCLUSIONS

In this paper, two motor imagery EEG classification methods are proposed to compare the results obtained with BCI2000 at the end of the testing paradigm. The pattern classification techniques, as described in this work, make possible the development of a motor imagery EEG signals analysis system which is accurate, simple and reliable enough to use in brain computer interface. We obtained better results when we used LDA classification methods that we implemented, compared to the results obtained with the BCI2000 software. In conclusion, the classifier used in BCI2000 can be improved to obtain better accuracy.

Future work will utilize the algorithms developed in this study, but the multiresolution wavelet analysis decomposition will be done with other types of mother wavelets.

REFERENCES

[1] S. Mason, G. E. Birch, "A general framework for brain-computer interface design", IEEE Trans. Neural Net. Syst. Rehab. Eng, 11, pp. 70.85, 2003.

[2] D. Millett, H. Berger., "From psychic energy to the EEG", Perspectives in biology and medicine, Vol. 44, 4, pp.522-542, 2001

[3] G. Pfurtscheller, E. Niedermeyer, F. H. Lopes da Silva, "EEG event-related desynchronization (ERD) and event-related synchronization (ERS)". Electroencephalography: Basic Principles, Clinical Applications and Related Fields, 4th Edition, Williams and Wilkins, Baltimore, 1998, pp 958-967.

[4] G. X. Bao, G. Song, "Pattern recognition of motor imagery EEG using wavelet transform", J. Biomedical Science and Engineering, 2008, 1, 64, 67

[5] H. Jasper, "The ten-twenty electrode system of the international Federation". Electroencephalogr. Clin. Neurophysiol., 10, pp: 370-375, 1958.

[6] R. Aldea, O. Eva, "Detecting sensorimotor rhythms from the EEG signals using the independent component analysis and the coefficient of determination," Proceedings of the 11th IEEE International Symposium on Signals Circuits and Systems (ISSCS'13), 11, pp. 13–16, 2013.

[7] R. Aldea, „Multiresolution wavelet analysis and Hurst estimation used for highlighting sensorimotor rhythms", Buletinul Institutului Politehnic Iasi, Vol. LIX (LXIII), pp. 63-71, 2013.

[8] D. Tărniceriu., "Bazele prelucrării numerice a semnalelor", Ed. Politehnium, Iaşi, 2008.

[9] I. Daubechies, "Ten Lectures on Wavelets", SIAM: Society for Industrial and Applied Mathematics; 1 edition (May 1992)

[10] R. Duda., P. Hart and D. Stork, "Pattern Classification" (2nd Edition), Wiley-Interscience, 2000.

Breast Cancer Diagnosis using Artificial Neural Networks with Extreme Learning Techniques

Chandra Prasetyo Utomo
YARSI E-Health Research Center
Faculty of Information Technology
YARSI University,
Jakarta, Indonesia

Aan Kardiana
YARSI E-Health Research Center
Faculty of Information Technology
YARSI University,
Jakarta, Indonesia

Rika Yuliwulandari
YARSI Genetics Research Center
Faculty of Medicine
YARSI University, Jakarta,
Indonesia

***Abstract*—Breast cancer is the second cause of dead among women. Early detection followed by appropriate cancer treatment can reduce the deadly risk. Medical professionals can make mistakes while identifying a disease. The help of technology such as data mining and machine learning can substantially improve the diagnosis accuracy. Artificial Neural Networks (ANN) has been widely used in intelligent breast cancer diagnosis. However, the standard Gradient-Based Back Propagation Artificial Neural Networks (BP ANN) has some limitations. There are parameters to be set in the beginning, long time for training process, and possibility to be trapped in local minima. In this research, we implemented ANN with extreme learning techniques for diagnosing breast cancer based on Breast Cancer Wisconsin Dataset. Results showed that Extreme Learning Machine Neural Networks (ELM ANN) has better generalization classifier model than BP ANN. The development of this technique is promising as intelligent component in medical decision support systems.**

Keywords—breast cancer; artificial neural networks; extreme learning machine; medical decision support systems

I. INTRODUCTION

The out of control development of cells in an organ is called tumors that can be cancerous. There are two kinds of tumors, benign and malignant. Benign or non-cancerous tumors are not spreading and are not life intimidating. In the other hand, malignant or cancerous tumor are expanding and life threatening [1]. Malignant breast cancer is defined when the growing cells are in the breast tissue. Breast cancer is the second overall cause of mortality among women and the first cause of dead among them between 40 and 55 ages [2].

Regular breast cancer diagnosis followed by appropriate cancer treatment can reduce the unwilling risk. It is suggested to do tumor evaluation test every 4-6 weeks. Based on that reason, benign and malignant detection based on classification features become very important [3].

Careful diagnosis in early detection has been proven to lessen the dead rate because of breast cancer [4]. Depend on the expertise, mistakes can be made by medical professionals while identifying a disease. With the help of technology such as data mining and machine learning, diagnosis can be more accurate (91.1%) when related to a diagnosis made by an experienced doctor (79.9%) [5].

This research is funded by YARSI University in the Internal Research Grant Schema 2013/2014.

ANN is one of the best artificial intelligence techniques for common data mining tasks, such classification and regression problems. A lot of research showed that ANN delivered good accuracy in breast cancer diagnosis. However, this method has several limitations. First, ANN has some parameters to be tuned in the beginning of training process such as number of hidden layer and hidden nodes, learning rates, and activation function. Second, it takes long time for training process due to complex architecture and parameters update process in each iteration that need expensive computational cost. Third, it can be trapped to local minima so that the optimal performance cannot be guaranteed. Numerous efforts had been attempted to get the solutions of neural networks limitations. Huang and Babri [6] proved that Single Hidden Layer Neural Networks (SFLN) with tree steps extreme learning process that called ELM can solve that problems.

In this paper, we revealed the implementation of artificial neural networks with extreme learning techniques in breast cancer diagnosis. The dataset used for experiments was Breast Cancer Wisconsin Dataset that was obtained from the University of Wisconsin Hospital, Madison from Dr. William H. Wolberg [7]. We compared the perfomance of ELM with conventional BP ANN with gradient descent based learning algorithms. Sensitivity, specificity, and accuracy were used as performance measurements. Results showed that ELM ANN generally produced better result than BP ANN.

The rest of this paper is organized as the following. Section 2 is dedicated as literature review. In this section, brief review of previous works in breast cancer diagnosis are presented. In Section 3, the concept, mathematical model, and training process of extreme learning machine are explained. In Section 4, experiments, results, and analysis are provided. Finally, conclusions and future works are given in Section 5.

II. LITREATURE REVIEW

The uses of classification systems in medical diagnosis, including breast cancer diagnosis, are growing rapidly. Evaluation and decision making process from expert medical diagnosis is key important factor. However, intelligent classification algorithm may help doctor especially in minimizing error from unexperienced practitioners [3].

Several techniques have been deployed to predict and recognize meaningful pattern for breast cancer diagnosis. Ryua

[8] developed data classification method, called isotonic separation. The performances were compared against support vector machines, learning vector quantization, decision tree induction, and other methods based on two-breast cancer data set, sufficient and insufficient data. The experiment results demonstrated that isotonic separation was a practical tool for classification in the medical domain.

Hybrid machine learning method was applied by Sahan [9] in diagnosing breast cancer. The method hybridized a fuzzy-artificial immune system with k-nearest neighbour algorithm. The hybrid method delivered good accuray in Wisconsin Breast Cancer Dataset (WBCD). They believe it can also be tested in other breast cancer diagnosis problems.

Comprehensive view of automated diagnostic systems implementation for breast cancer detection was provided by Ubeyli [10]. It compared the performances of multilayer perceptron neural network (MLPNN), combined neural network (CNN), probabilistic neural network (PNN), recurrent neural network (RNN) and support vector machine (SVM). The aim of that works was to be a guide for a reader who wants to develop this kind of systems.

Numerous combinations and hybrid systems used neural networks as a component. However, since almost all of the employed neural networks are conventional gradient descent BP ANN, the novel or hybrid method still suffered the neural networks drawbacks that were mentioned in the previous section.

III. Extreme Learning Machine

Huge efforts had been attempted to solve the weaknesses of BP ANN. Huang and Babri [6] demonstrated that single hidden layer feedforward neural networks (SLFN) with at most m hidden nodes was capable to estimate function for m different vectors in training dataset.

Given m instances in $\mathbf{D} = \{(\mathbf{x}^{(k)}, \mathbf{t}^{(k)}) \mid \mathbf{x}^{(k)} \in \mathbf{R}^n, \mathbf{t}^{(k)} \in \mathbf{R}^p, k = 1,..,m\}$ as training dataset where $\mathbf{x}^{(k)} = [x_1^{(k)}, x_2^{(k)}, \ldots, x_n^{(k)}]^T$ as features and $\mathbf{t}^{(k)} = [t_1^{(k)}, t_2^{(k)}, \ldots, t_p^{(k)}]^T$ as target. A SLFN with M number of hidden nodes, activation function $g(x)$ in hidden nodes, and linear activation function in output nodes is mathematically wrote as:

$$\sum_{i=1}^{M} \beta_i g_i(\mathbf{x}^{(k)}) = \sum_{i=1}^{M} \beta_i g(\mathbf{w}_i \cdot \mathbf{x}^{(k)} + b_i) = \mathbf{o}^{(k)}, \quad k = 1, \ldots, m \tag{1}$$

where

$\mathbf{w}_i \in \mathbf{R}^n$ is the weights between the input nodes and the i-th hidden node

$$\mathbf{w}_i = [w_{i1}, w_{i1}, \ldots, w_{in}]^T, \tag{2}$$

$\boldsymbol{\beta}_i \in \mathbf{R}^p$ is the weights between the i-th hidden node and the output nodes

$$\boldsymbol{\beta}_i = [\beta_{i1}, \beta_{i2}, \ldots, \beta_{ip}]^T, \tag{3}$$

$\mathbf{w}_i \cdot \mathbf{x}^{(k)}$ is the inner product of $\mathbf{w}_i$ and $\mathbf{x}^{(k)}$,
b_i is the bias of the i-th hidden node,
$\mathbf{o}^{(k)} \in \mathbf{R}^p$ is the output of neural network for k-th vector.

SLFN can approximate m vectors means that there are exist $\mathbf{w}_i$, $\boldsymbol{\beta}_i$, and b_i, such that:

$$|| \mathbf{o}^{(k)} - \mathbf{t}^{(k)} || = 0 \tag{4}$$

$$\sum_{i=1}^{M} \beta_i g(\mathbf{w}_i \cdot \mathbf{x}^{(k)} + b_i) = \mathbf{t}^{(k)}, \quad k = 1, \ldots, m \tag{5}$$

Equation (5) can be written as:

$$\mathbf{H}\boldsymbol{\beta} = \mathbf{T}, \tag{6}$$

where

$\mathbf{H} \in \mathbf{R}^{m \times M}$ is the hidden layer output matrix of the neural networks.

$$\mathbf{H} = \begin{bmatrix} g(w_1 \bullet x^{(1)} + b_1) & \cdots & g(w_M \bullet x^{(1)} + b_M) \\ \vdots & \ddots & \vdots \\ g(w_1 \bullet x^{(m)} + b_1) & \cdots & g(w_M \bullet x^{(m)} + b_M) \end{bmatrix} \tag{7}$$

$\boldsymbol{\beta} \in \mathbf{R}^{M \times p}$ is the weights between hidden and output layers

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_M^T \end{bmatrix}, \tag{8}$$

$\mathbf{T} \in \mathbf{R}^{m \times p}$ is the target values of m vectors in training dataset

$$\mathbf{T} = \begin{bmatrix} t^{(1)T} \\ \vdots \\ t^{(m)T} \end{bmatrix}, \tag{9}$$

In the traditional gradient descent based learning algorithm, weights $\mathbf{w}_i$ which was connecting the input layer and hidden layer and biases b_i in the hidden nodes were needed to be initialized and tuned in every iteration. This was the main factor which often made training process of neural became time consuming and the trained model may not reach global minima.

Huang [11] proposed minimum norm least-squares solution of SLFN which didn't need to tune those parameters. Training SLFN with fixed input weights $\mathbf{w}_i$ and the hidden layer biases b_i was similar to find a least square solution $\widehat{\boldsymbol{\beta}}$ of the linear system $\mathbf{H}\boldsymbol{\beta} = \mathbf{T}$:

$$\left\|\mathbf{H}(\mathbf{w}_1, \dots, \mathbf{w}_\mathrm{M}, \boldsymbol{b}_1, \dots, \boldsymbol{b}_\mathrm{M},)\widehat{\boldsymbol{\beta}} - \mathbf{T}\right\| = \min_{\boldsymbol{\beta}} \|\mathbf{H}(\mathbf{w}_1, \dots, \mathbf{w}_\mathrm{M}, \boldsymbol{b}_1, \dots, \boldsymbol{b}_\mathrm{M},)\boldsymbol{\beta} - \mathbf{T}\|. \quad (10)$$

The smallest norm least squares solution of that linear system was

$$\widehat{\boldsymbol{\beta}} = \mathbf{H}^{\dagger}\mathbf{T} \quad (11)$$

where $\mathbf{H}^{\dagger}$ was the *Moore-Penrose generalized inverse* of matrix H. This solution had three important properties which were minimum training error, smallest norm of weights, and unique solution which is $\widehat{\boldsymbol{\beta}} = \mathbf{H}^{\dagger}\mathbf{T}$.

The above minimum norm least-square solution for SLFN was called extreme learning machine (ELM). Given *m* instances in training dataset $\mathbf{D} = \{(\mathbf{x}^{(k)}, \mathbf{t}^{(k)}) \mid \mathbf{x}^{(k)} \in \mathbf{R}^n, \mathbf{t}^{(k)} \in \mathbf{R}^p, k = 1,..,m\}$, activation function $g(x)$, and number of hidden node *M*. The training process of ELM is the the following:

1. Randomly set input-hidden layer weights $\mathbf{w}_i$ and bias b_i, $i = 1,\dots,M$.
2. Compute the matrix of hidden layer output **H**
3. Compute the hidden-output layer weights $\widehat{\boldsymbol{\beta}}$ for $\widehat{\boldsymbol{\beta}} = \mathbf{H}^{\dagger}\mathbf{T}$ where $\mathbf{T} = [\mathbf{t}^{(1)},\dots, \mathbf{t}^{(m)}]$.

Based on that definition, there are three main differences between BP ANN and ELM ANN. First, BP ANN needs to tuning several parameters, such as number of hidden nodes, learning rates, momentum, and termination criteria. On the other hand, ELM ANN is a simple tuning free algorithm. The only one to be defined is number of hidden nodes. Second, BP ANN works only for differentiable activation functions in hidden and output nodes while ELM ANN can use both differentiable and undifferentiable activation functions.

Finally, BP ANN get trained model which has minimum training error so that there is a possibility to finish in local minima. On the other hand, ELM ANN get trained model which has minimum training error and smallest norm of weight so that it can produce better generalization model and reach global minima [12].

IV. Experiments, Results, and Analysis

This section discussed about experimental design, generated results, and analytical process in order to get valid conclusion.

A. Experiments

The experiments consisted of three main steps, which were data gathering, data preprocessing, and performance evaluating. The dataset used in this experiment was Breast Cancer Wisconsin Dataset obtained from the University of Wisconsin Hospital, Madison from Dr. William H. Wolberg [7]. The data has 699 instances with 10 attributes plus the class attributes. The class distribution are 65.5% (458 instances) for benign and 34.5% (241 instances) for malignant. The attribute information can be seen in TABLE I.

TABLE I. Attribute Information

#	Attributes	Domain
1	Sample Code Number	id number
2	Clump Thickness	1 – 10
3	Uniformity of Cell Size	1 – 10
4	Uniformity of Cell Shape	1 – 10
5	Marginal Adhesion	1 – 10
6	Single Epithelial Cell Size	1 – 10
7	Bare Nuclei	1 – 10
8	Bland Chromatin	1 – 10
9	Normal Nucleoli	1 – 10
10	Mitoses	1 – 10
11	Class	2: benign; 4: malignant

In the second step, the raw dataset was preprocessed to produce well-from data that suitable for training and testing process. The first attribute, sample code number, was removed because it was not relevant to the diagnosis. The next nine attributes were normalized into [-1, 1] and used as predictor. The last attribute was transformed to 0 (benign) and 1 (malignant) such that it can be properly fitted to the standard BP ANN and ELM ANN implementation.

The method in this experiment was k-fold crossvalidation with k = 5. This means, the data were randomly divided into 5 partitions. There were 5 experiments. In the each experiment, a partition was used as testing data and the rest partitions were treated as training data.

The standard performance measurement for classification problem was accuracy. However, since the class distribution was not balanced, it was important to use specificity and sensitivity as supplementary measurements. In addition, to minimize the effect of random generated weights in BP ANN and ELM ANN, each experiment was run three times and the average results were noted.

$$accuracy = \frac{True\ Positve + True\ Negative}{All\ Testing\ Data} \quad (13)$$

$$sensitivity = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (14)$$

$$specificity = \frac{True\ Negative}{False\ Positive + True\ Negative} \quad (15)$$

B. Results and Analysis

With 5-fold crossvalidation method and each experiment were run three times, there will be 15 experiments in total. The whole steps had been done in computer with Intel® Core™ i3, 4096MB RAM, and Windows 7 OS. The results of ELM are given in TABEL II.

TABLE II. ELM PERFORMANCES

Experiments	Running	Sensitivity	Specificity	Accuracy
#1	1	0,902	0,965	0,942
	2	0,882	0,977	0,942
	3	0,882	0,965	0,934
#2	1	0,975	0,958	0,963
	2	0,950	0,979	0,971
	3	1,000	0,969	0,978
#3	1	0,909	0,988	0,956
	2	0,964	0,963	0,963
	3	0,964	0,975	0,971
#4	1	0,957	1,000	0,985
	2	0,957	0,989	0,978
	3	0,957	1,000	0,985
#5	1	0,957	0,967	0,964
	2	1,000	0,956	0,971
	3	0,957	0,956	0,956

In order to compare the ELM ANN performances, training and testing with BP ANN were conducted with identical experimental design. The performances of BP ANN can be seen in TABLE III.

TABLE III. BP ANN PERFORMANCES

Experiments	Running	Sensitivity	Specificity	Accuracy
#1	1	0,875	0,975	0,934
	2	0,750	0,988	0,891
	3	0,804	0,963	0,898
#2	1	0,855	0,987	0,927
	2	0,871	0,973	0,927
	3	0,903	0,973	0,941
#3	1	0,825	0,949	0,897
	2	0,860	0,987	0,934
	3	0,825	0,975	0,912
#4	1	0,790	0,975	0,898
	2	0,825	0,975	0,912
	3	0,807	0,975	0,905
#5	1	0,892	1,000	0,949
	2	0,892	1,000	0,949
	3	0,877	1,000	0,942

To have clear view between the performances of ELM ANN and BP ANN, the results were transformed to graphical charts. In each performance measurement, the average values were computed in each experiment. Fig 1 shows the comparison of average sensitivity rates between BP ANN and ELM ANN. The comparison of specificity rates were given in Fig 2 while accuracy rates can be seen in Fig 3.

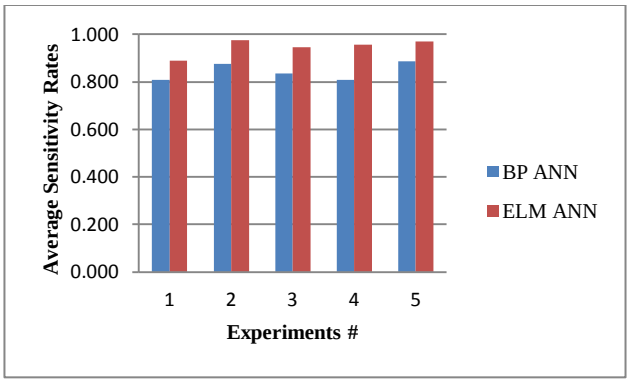


Fig. 1. Average sensitivity rates of BP ANN dan ELM

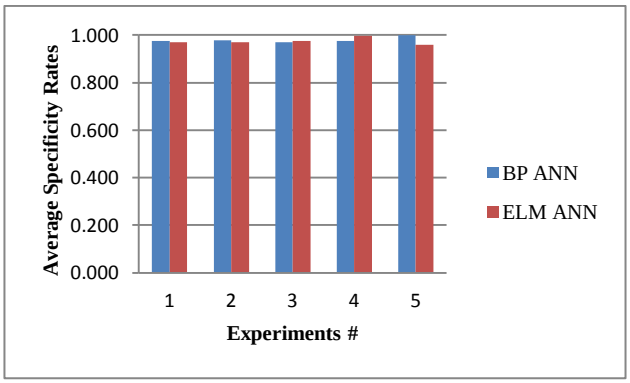


Fig. 2. Average specificity rates of BP ANN dan ELM

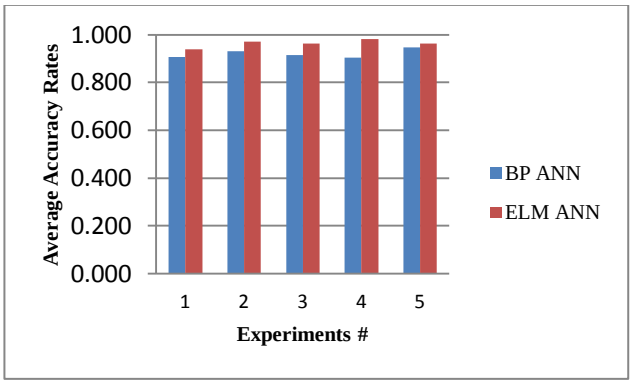


Fig. 3. Average accuracy rates of BP ANN dan ELM

Based on Fig 1 and Fig 3, we can see that ELM ANN was superior compared to BP ANN. ELM ANN has better performances in term of sensitivity and accuracy in all experiments.

However, in term of specificity, BP ANN has better performance in three experiments which were experiment #1, #2, and #5.

To get general conclusion, overall comparison need to be computed. In each performance measurement, commulative average rates were matched. Fig 4 displays the whole sensitivity, specificity, and accuracy average rates between BP ANN and ELM ANN. Result showed that, generally ELM ANN were better than BP ANN.

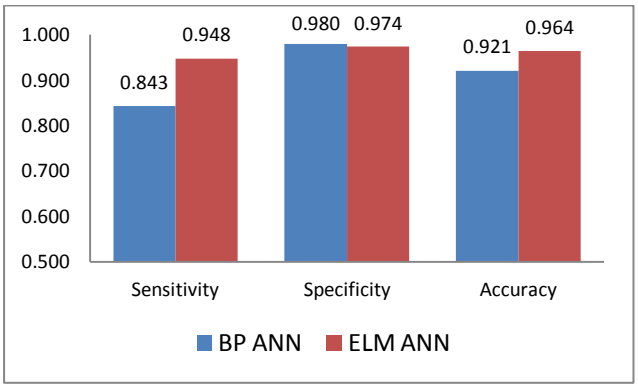


Fig. 4. Overall average rates between BP ANN dan ELM

V. Conlusion and Future Works

The performances of ELM ANN were generally better than BP ANN in breast cancer diagnosis. Although the specificity rate was slightly lower than BP ANN, it can be clearly seen that ELM ANN remarkably improved the sensitivity and accuracy rates. Based on these results, we can conclude that ELM ANN has better generalization model than BP ANN in diagnosing breast cancer based on Breast Cancer Wisconsin Dataset.

There are some necessary works to be done in near future. First, it is important to communicate the model with domain expert. The hybrid of ELM ANN with Decision Tree or any other technique that can produce meaningful knowledge representation will be promising. Second, to make intelligent diagnosis tool that can be used by end user, it is necessary to develop interactive user interface.

The development of interfaces in mobile, desktop, or web application may be useful. Third, there are new cases added regularly in the hospital. Developing intelligent diagnosis systems that can not only learn from available data in repositories but also from newly available data will be required.

References

[1] T. S. Subashini, V. Ramalingam, and S. Palanivel, "Breast mass classification based on cytological patterns using RBFNN and SVM," *Expert Syst. Appl.*, vol. 36, no. 3, pp. 5284–5290, Apr. 2009.

[2] TheAmericanCancerSociety, "What is Breast Cancer." .

[3] M. F. Akay, "Support vector machines combined with feature selection for breast cancer diagnosis," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 3240–3247, Mar. 2009.

[4] D. West, P. Mangiameli, R. Rampal, and V. West, "Ensemble strategies for a medical diagnostic decision support system: A breast cancer diagnosis application," *Eur. J. Oper. Res.*, vol. 162, no. 2, pp. 532–551, Apr. 2005.

[5] R. W. Brause, "Medical Analysis and Diagnosis by Neural Networks," in *Proceeding ISMDA '01 Proceedings of the Second International Symposium on Medical Data Analysis*, 2001, pp. 1–13.

[6] G. B. Huang and H. A. Babri, "Upper bounds on the number of hidden neurons in feedforward networks with arbitrary bounded nonlinear activation functions.," *IEEE Trans. Neural Netw.*, vol. 9, no. 1, pp. 224–9, Jan. 1998.

[7] W. H. Wolberg and O. L. Mangasarian, "Multisurface method of pattern separation for medical diagnosis applied to breast cytology," in *Proceedings of the National Academy of Sciences, U.S.A., Volume 87, December 1990*, pp. 9193–9196.

[8] Y. U. Ryu, R. Chandrasekaran, and V. S. Jacob, "Breast cancer prediction using the isotonic separation technique," *Eur. J. Oper. Res.*, vol. 181, no. 2, pp. 842–854, Sep. 2007.

[9] S. Sahan, K. Polat, H. Kodaz, and S. Güneş, "A new hybrid method based on fuzzy-artificial immune system and k-nn algorithm for breast cancer diagnosis.," *Comput. Biol. Med.*, vol. 37, no. 3, pp. 415–23, Mar. 2007.

[10] E. D. Übeyli, "Implementing automated diagnostic systems for breast cancer detection," *Expert Syst. Appl.*, vol. 33, no. 4, pp. 1054–1062, Nov. 2007.

[11] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, no. 1–3, pp. 489–501, Dec. 2006.

[12] C. P. Utomo, "The Hybrid of Classification Tree and Extreme Learning Machine for Permeability Prediction in Oil Reservoir," *Int. J. Comput. Sci. Issues*, vol. 10, no. 1, pp. 52–60, 2013.