A hybrid model combining neural networks and decision tree for comprehension detection.

James O'Shea[1], Keeley Crockett[1], Wasiq Khan[1], Zuhair Bandar[2]
[1]School of Computing, Mathematics and Digital Technology
Manchester Metropolitan University
Chester Street, Manchester, M1 5GD, UK
[2]Silent Talker Ltd, Manchester, UK
J.D.OShea@mmu.ac.uk

Abstract— **The Artificial Neural Network is generally considered to be an effective classifier, but also a "Black Box" component whose internal behavior cannot be understood by human users. This lack of transparency forms a barrier to acceptance in high-stakes applications by the general public. This paper investigates the use of a hybrid model comprising multiple artificial neural networks with a final C4.5 decision tree classifier to investigate the potential of explaining the classification decision through production rules. Two large datasets collected from comprehension studies are used to investigate the value of the C4.5 decision tree as the overall comprehension classifier in terms of accuracy and decision transparency. Empirical trials show that higher accuracies are achieved through using a decision tree classifier, but the significant tree size questions the rule transparency to a human.**

Keywords—knowledge rule extraction, artificial neural networks, decision trees, backpropagation, comprehension, FATHOM, Silent Talker, non-verbal behavior.

I. INTRODUCTION

Non-Verbal Behaviour (NVB) was first studied systematically by Charles Darwin and it has become a well-established part of sciences such as biology and psychology. NVB consists of all of the signs and signals - visual, audio, tactile and chemical used by human beings to express themselves apart from speech and manual sign language. It has been postulated that NVB features are indicators of internal mental states, in particular that they can be used to detect deception during interviews [1]. The first system to classify deceptive behaviour automatically, Silent Talker (ST), used Artificial Neural Networks [2].

The Silent Talker architecture is highly flexible, and has been adapted to monitor human comprehension in clinical trials using non-verbal behaviour, employing ANN classifiers [3]. This version is known as FATHOM and currently work is underway to incorporate FATHOM in an intelligent tutoring system to provide round-the-clock support in the form of learner-adaptive online teaching and learning tutorials. For both FATHOM and ST there has been great interest in how the system works i.e. which non-verbal indicators are actually conveying the information to perform the classification. This is particularly true for Lie Detection where interrogators are looking for techniques they can apply during questioning and suspects are looking for countermeasures they can use to avoid being detected, for example, the well-known myth that looking up and to the right indicates lying. Unfortunately, although ANNs are powerful and versatile components in the AI toolbox, they are also black boxes with no ready explanations of how they achieve their ends and this has been a concern for decades [4].

There are many other fields than education in which ANNs may make high-stakes decisions and some progress has been made in extracting rules from ANNs, although the degree to which solutions to reasonably complex problems could be understood by a non-AI specialist remains debatable. These include classifying incipient faults in a power transformer [5], hydrological modelling [6], Credit-Risk Evaluation [7] and software cost estimation [8]. Some progress has been made in extracting rules from recurrent neural networks by transforming them to finite state machines [9], and [10] has attempted to unify various neuro-fuzzy rule approaches for ruled generation from recurrent and feedforward neural networks in a single soft computing framework. Nevertheless, in analysing a study of using neural networks to predict academic performance of college students one year in advance, Schneider et al. [11] observed that the basic problems of communicating how they reach their conclusions in meaningful terms has yet to be solved. They highlighted the problem of explaining how a combination of currently high subject performances could lead to an anticipated decrease in the student's achievement.

Decision trees [12] are highly effective for classification tasks. They are also considered inherently transparent in explaining how they reach their conclusions and may be expressed in the form of production rules, which are generated, by learning and reasoning from feature-based examples. Many studies have been conducted to compare decision trees with neural networks – a more recent study of multiple classifiers can be found in Delgado et al. [13]. In general, ANNs take longer to train than decision trees due to the large number of iterations required to ensure training reaches its full potential [14]. Classification accuracy is largely dependent on the dataset, but the transparent nature of decision trees gives insight into the relationships between features [14]. In such a domain as the analysis of NVB for comprehension detection, decisions trees would provide an insight into key behaviours and their interactions.

In the FATHOM architecture to date, classification of comprehension / non-comprehension has been performed by a

single, final back propagation artificial neural network (BPANN), preceded by layers of BPANNs that process individual features. It is this final stage in which an intervention should be possible to explain how these features indicate comprehension / non-comprehension. Therefore, the research questions addressed in the work presented in this paper are:

1. Can the final ANN classifier be replaced by a decision tree without loss of performance?

2. Can the decision tree be converted into comprehensible production rules?

For a comprehensible rule set be possible, there must be a limited number of rules for the human user to interpret and these are proportional to the number of nodes in the tree. Consequently, the primary interest in answering question 2 is whether or not the tree has a manageable number of nodes.

To answer these research questions, two datasets collected from FATHOM studies have been used. The experimental study known as "Termites", reported in [15] was used to identify whether high and low human comprehension associated multi-channels of non-verbal behaviour reside within a video-recorded British (UK-based/English speaking) sample of participants in a class room environment. The Termites exploratory study builds upon lessons learned in prior work [3] where evidence was found that comprehension / non-comprehension could be detected in an African female population sample using a BPANN. This second study is known as HIV Informed Consent.

This paper continues as follows: Section II reviews related work in non-verbal behaviour and comprehension, and then describes the FATHOM comprehension monitoring system that uses BPANNs. Section III describes the comprehension scenarios from which the two datasets used in this study were obtained. Section IV and V describe the experimental methodology and results. Conclusions and recommendations for future work are presented in section VI.

II. RELATED WORK

A. Non-verbal Behaviour and Comprehnsion

Non-verbal behaviour comprises all of the signals or cues, which human beings use to communicate, including visual, audio, tactile and chemical components [16, 17]. During a spoken dialogue, humans will often transmit non-verbal cues before the verbal component [17], which can be used to detect the sender's state. It has been recognised that the face is a source of rich information in terms of exhibiting meaningful non-verbal behaviour. Little work has been done in the automatic detection of classification of non-verbal behaviour. Traditional methods employed human judges to code each channel [18, 19]. However, each judge needs to be trained and will provide a subjective opinion on the behaviour being delivered by a particular channel. The process is time consuming and an impossible task for a human to monitor more than a limited number of channels accurately.

Two recent research strategies for acquiring non-verbal behavioural cues have attracted attention in the literature; these are Facial Microexpressions [1] and using the Microsoft Kinect computer vision algorithm [20]. Micro-expressions are said to be a small "universal" set of expressions of extreme emotion: disgust, anger, fear, sadness, happiness, surprise, and contempt, and a formalised method of encoding them was defined by Ekman. The weaknesses of this technique are: its results are largely based on highly artificial "posed" images using actors or students provided with highly specific instructions [21,22] or even training in how to produce facial actions [23,24], low numbers of detectable Ekman micro-expressions in spontaneous interviews [25] and a low Classification Accuracy (CA) for those micro-expressions actually found [26].

The Microsoft Kinect is primarily aimed at observing whole body gestures in commercial video game applications. However, there has been some interest in adapting it for NVB research. For example, facial expressions have been investigated as indicators of happiness, anger, sadness and surprise that are integrated with the head pose changing information to conceive the human interaction with 3D sensing technology [27]. Although it should be noted that the experimental results show emotional and head position change instead of discrete level accuracy in terms of emotional classification for the four aforementioned emotions. Likewise, it was tested on a limited participants (i.e. 20) as well as insufficient facial channels (i.e. 12). Typically, psychological experiments do not provide any methodology for applying these population sample differences to classify particular individuals.

FATHOM (described in Section II, C) is distinguished from these two techniques in three respects. Firstly, it uses large numbers of features at a much finer level of granularity than body gestures or facial expressions. Secondly, the domain it operates in, human comprehension has not been a previous subject of AI research. Thirdly, it does classify an individual person's state of comprehension / non-comprehension based the non-verbal behaviour. Fathom does not rely on high frame-rate cameras or constrained recording environments that facilitate the setup of the technology, nor does it depend on specialised hardware whose future availability may be dependent on market forces (such as the game-oriented Kinect) – making it suited for everyday classroom use.

B. Non-Comprehension

Non-comprehension is regarded as "a state of knowledge that ranges from uncertainty to complete lack of understanding of the materials under discussion" [28], i.e. an absence of comprehension. The vast majority of research on comprehension concerns reading and the understanding of written text, initially by identifying the main ideas in the text [28, 29]. A further elaboration is the view that successful comprehension depends on the construction of a coherent representation of text in memory [30]. Despite the traditional bias towards reading texts, there has been interest for some time in comprehending audio and video materials in language teaching [31] and the informed consent process [32]. At a more abstract level, the comprehension of metaphors, requires

thinking beyond the literal meaning in order to understand the figurative meaning of the sentence [33] – yet metaphors and similes are frequently used by good teachers to convey complex ideas. In the completely independent field of advertising, a controlled degree of cognitive complexity is considered desirable, where confronting an audience with a cognitive challenge generates an appreciative payoff if they can solve the challenge [34]. So the non-comprehension state may be characterized as an inability to extract and characterize the salient elements of information received, an inability to model such information in a more abstract form or an inability to generalize from a specific meaning to more abstract thoughts about such a communication.

C. FATHOM

FATHOM utilises a bank of BPANN's to capture, monitor and detect multiple channels of human non-verbal behaviour continuously. FATHOM has been successfully shown to detect non-verbal behaviour associated with comprehension in two studies [3] [15].

Input to FATHOM is currently offline through recorded videos, which are streamed into FATHOM where a series of BPANN facial object locators, identify the location in a video frame of key visual features such as the eyes. For each non-verbal behavioural feature identified from a specific visual feature, the BPANN facial object pattern detectors identify its state i.e. the left eye is half-open. The NVBs identified are then coded into individual channels and group channels i.e. all channels associated with eye behaviour.

States are typically collated over a time interval, e.g. 3 seconds grouped into one vector for Silent Talker– but this can be varied depending on the problem domain and FATHOM uses a 1-second interval. Classification features (patterns) are extracted from aggregated video-streamed frames over the time interval and compiled to form a vector. Each vector is passed to the final BPANN Comprehension classifier which outputs a value between +1 and -1, indicating whether the person exhibits high comprehension (+1) or low comprehension (-1) during that period of time. If there were insufficient information in the vector during a specific time slot, FATHOM would recognise this and categorise the timeslot as unclassifiable. At the end of a session i.e. a tutorial, the overall comprehension/non-comprehension classification level is displayed.

FATHOM simultaneously monitors 40 non-verbal behavioural channels that include 20 channels capturing facial features such as blushing and 16 channels capturing eye movement such as right eye looking left. An overview of the FATHOM architecture can be seen in Figure 1.

The work presented in this paper investigates the consequences of replacing the BPANN comprehension classifier in the FATHOM system by a C4.5 decision tree [12], to answer questions about their relative performance and transparency.
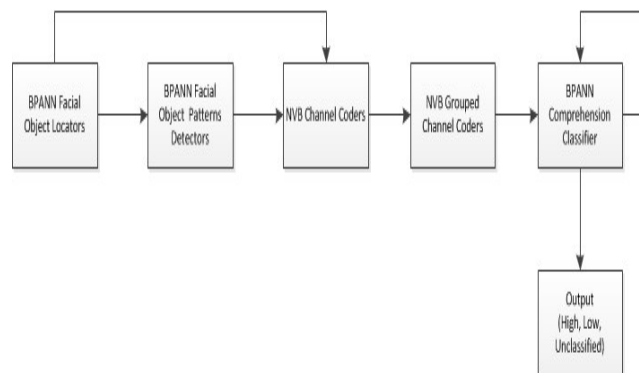

Fig.1 FATHOM Architecture

III. COMPREHENSION SCENARIOS

This section outlines the two comprehension scenarios used to collect the data.

A. Study 1:HIV Informed Consent

The first comprehension study was undertaken in Tanzania in Africa by FHI-360 [35] in collaboration with the National Institute for Medical Research (NIMR) [36]. NIMR enlisted sexually active women aged 18-35, who were native Kiswahili-speakers. 292 participants took part in the study. Two different experimental conditions (tasks) were used for data collection: condition A was designed to be familiar and easy-to-comprehend (condom use) and condition B was designed to be unfamiliar and intentionally hard-to-comprehend (the effects of HIV viral mutation on antiretroviral treatment). Each participant listened to a short learning task script and then received the associated ten closed and open-ended questions with randomisation applied. Task order was also randomised so that half of the participants completed task A followed by task B and vice versa.

B. Study 2: Termites

Prior to the study a short learning topic was selected, which was a factual digital video on Termites with a total duration of 8 minutes 40 seconds. The Termite video was targeted at the general public with no age restriction and covered: functional architectural aspects of the termite mounds, roles within the social structure of a termite colony and locations where termite colonies thrive. Two experts (Academic Professors in the field) on the subject area were recruited to develop ten difficult (hard) questions and ten easy questions related to the video content. The experts agreed both the question difficulty levels and the contents of the answer that the participants should provide. The experts were required to devise five open questions and closed questions within each set of hard and easy questions. At the same time, the experts noted down the correct answer(s) for each question, which were later incorporated into a scoring scheme.

Forty participants were selected to participate in the study, from academic and technical staff at the Manchester Metropolitan University (MMU) in the UK. The sample was composed of 20 males and 20 females. The males had a mean age of 41 years old (SD = 14 years) and the females had a mean age of 39 years old (SD = 14 years). Each participant was invited to engage individually in a short learning task, which was comprised of watching a short video on Termites and then answering a small set of associated assessment questions whilst being video recorded.

IV. EXPERIMENTAL METHODLOGY

The experimental methodology was to take the pair of datasets outlined in Section III and use them to train and evaluate C4.5 decision trees to replace the final stage BPANN classifier.

A. Using a Back Propagation ANN as the final classifier

For each study, FATHOM's object locators and pattern detectors were used to extract and collate the non-verbal vector-based dataset for the purpose of training the final BPANN classifier. For both studies, HIV Informed Consent and Termites, each vector in the final dataset covered a 1-second time period and represented the state changes for the compiled non-verbal channels over the period. Each channel was normalised in the range +1 to -1. The last attribute in each vector was the desired classification, with discrete values of +1 for comprehension and -1 for non-comprehension. The following training parameters (determined from previous exploratory cross-validation sessions) were used to train the single hidden layer neural network in the Fathom training application:

- Topology: 40:20:1
- Accept value: 1.0 (output >= 0.0 equals comprehension AND output <0.0 equals non-comprehension)
- Maximum epochs: 10,000
- Checking epochs: 250, i.e. at every 250th epoch the total Classification accuracy (CA) was checked and if there was no improvement training was terminated.
- Learning rate (η) was set at 0.005.
- Weight initialisation: automatic range ($0\pm1/\text{sqrt(fan-in)}$) where fan-in represents the number of inputs entering the neuron.
- Cross-validation: 10-folds

For study 1, eighty randomly selected participant videos (from the 292 obtained in the study) comprised the HIV Informed Consent dataset containing 71,787 vectors with 63.5% comprehension and 36.5% non-comprehension. For study 2, the forty participant videos yielded 16,951 comprehension vectors and 23,857 non-comprehension vectors. The study 2 Termites dataset was composed of 40,808 vectors with 41.5% in the comprehension class.

B. Using a C4.5 Decision Tree as the final classifier

In order to use a decision tree as a comprehension classifier the final, BPANN, comprehension classifier shown in Figure 1 was replaced by the C4.5 decision tree algorithm.

1) The experiment consisted of a series of trials using different degrees of pruning to find the optimal C4.5 decision trees and determine the extent to which they could be pruned. The Weka implementation of C4.5, known as J48 was used. This was achieved by establishing a baseline decision tree for each scenario setting the pruning parameters to Confidence Interval (CI) = 0.25 and minimum number of objects = 2 cases per leaf (i.e. the default settings)

2) This was followed by fixing the minimum number of objects (MNO) at 2 and conducting a series of trials over a range of confidence interval values to determine which provides the greatest improvement in CA over the baseline tree performance.

3) Then the complementary process was performed, fixing the CI at 0.25 and conducting a series of trials over a range of values of MNO.

4) Finally, further experiments were performed varying confidence interval and MNO independently, to find the most severely pruned tree for each dataset, which, was not significantly worse than the baseline in terms of CA.

The initial ranges used for the experiments were, for CI: 0.25, 0.2, 0.15, 0.1, 0.05, and for MNO: 2, 5, 10, 15, 20.

V. RESULTS

A. BPANN Comprehension Classifier

Table I shows the overall best BPANN Classifiers for both studies. Comprehension (C%) and Non-comprehension (NC%) are the percentages of comprehension and non-comprehension vectors, respectively, which were classified correctly. Overall % is the total normalised percentage of comprehension and non-comprehension vectors classified correctly.

TABLE I: BPANN RESULTS

	C%	NC%	Overall CA %
Study 1: HIV Informed Consent	88.08	86.79	**87.44**
Study 2: Termites	72.77	84.09	**78.43**

B. C4.5 Comprehension Classifier

Table II shows the results of varying the Confidence Interval used for pruning in decision tree construction (Pruning CI).

Pruning CI	0.25	0.20	0.15	0.10	0.05
CA%	94.56	94.46	94.40	94.34	94.16
Tree size	4567	4325	4387	4181	4039

The corrected re-sampled *t*-test was used for significance testing of the following results. A significant drop in CA was observed at a Pruning CI of 0.20. On investigating further, values of 0.24, 0.23, and 0.22 give classification accuracies of 94.53%, 94.51%, and 94.50% respectively with a significance drop occurring when Pruning CI is 0.22. None of the other pruning levels produces an increase in CA. Table III shows the results obtained from varying the minNumObj − the minimum number of instances per leaf.

TABLE III: STUDY 1: HIV INFORMED CONSENT STUDY: MINNUMOBJS

minNum Obj	2	5	10	15	20	25	30
CA%	94.56	93.26	91.40	89.82	88.64	87.67	86.88
Tree Size	4567	4121	3561	2557	1997	1821	1211

There was a significant drop in the overall accuracy when the minimum number of objects increased to 5. On further investigation when minNumObj were 3 and 4 the classification accuracies were 94.15 and 93.67 respectively. The tree size when minNumObj is 4 was 4161 leaf nodes.

TABLE IV: STUDY 2: TERMITES: PRUNING

Pruning CI	0.25	0.20	0.15	0.10	0.05
CA%	94.38	94.24	94.01	94.09	93.88
Tree size	3317	3169	3121	3047	2975

TABLE V: STUDY 2: TERMITES: MINNUMOBJS

minNum Obj	2	5	10	15	20	25	30
CA%	94.38	92.60	90.07	88.05	86.44	85.02	84.14
Tree size	3317	2605	1895	1579	1299	1171	1015

Using a similar pruning strategy for Termites as that for HIV Informed Consent resulted in a DT of 2992 nodes.

C. Discussion

The findings support a positive answer to the first question, "Can the final ANN classifier be replaced by a decision tree without loss of performance?" Classification accuracies of 94.56% for HIV Informed Consent and 94.38% for Termites are substantially higher than their respective equivalents using the BPANN (87.44% and 78.43% respectively). However, it could be argued that the decision trees are over-trained, at 4567 nodes for the HIV Informed Consent dataset and 3317 nodes for Termites. Pruning the trees leads to reductions to sizes of 4161 and 2992 nodes respectively, without a significant

reduction in CA. Also a further experiment showed that an over-trained tree (rote-learning of the dataset) for the Termites dataset has a size of 6489 nodes; more than double the size of the optimally pruned tree.

In answer to the second question, "Can the decision tree be converted into comprehensible production rules?", the trees are clearly too large to produce graphs or production rules that could be readily understood by a human.

A further interesting question emerged, "How severely may DTs be pruned before they drop below the performance of the corresponding BPANN?" The HIV Informed Consent CA of 86.88% for (1211 nodes) is close to the best BPANN CA of 87.44%. This provides further evidence that DTs are effective at modelling the comprehension domain, but take us little closer to the ideal of producing a tree that is readily comprehensible by a human.

VI. CONCLUSIONS AND FUTURE WORK

The evidence from this study supports the view that complex internal mental states associated with comprehension are classifiable through external non-verbal behaviors. It also provides evidence that machine (AI) based classifiers of non-verbal behavior are the best way forward.

However, because of the large number of production rules that would be generated from such large trees as produced in these experiments, the findings do not support the view that such behaviors can be easily codified into rule-of-thumb heuristics which humans could apply to perform such classifications for themselves, in real time. Consequently, neither does it support their uptake by a human population who need some sort of understandable verification of AI components used in high-stakes applications.

We propose that future work should explore several options to simplify the decision trees and their representation. First, pre-preprocessing the data to cleanse it, particularly removing outliers, noise and conflicting records - all of which might be better handled by the BPANN than DTs. Second, investigation of the potential to reduce the number of input channels – through empirical experiment, by identifying the potentially lowest contributing channels through calculating information content and by grouping channels. Third, by using fuzzy rule extraction or random forest techniques to reduce the rule sets extracted from the more efficient trees to a more tractable size. These approaches should contribute to a move to AI classifiers, which are themselves comprehensible by the human population.

REFERENCES

[1] Ekman training website. [online] Available: http://www.paulekman.com/about-paul-ekman-group-llc/ [Accessed 31/1/2018]

[2] Rothwell, J. Bandar, Z. O'Shea, J. McLean, d. "Charting the behavioural state of a person using a backpropagation neural network". Neural Computing & Applications, vol. 16, pp. 327-339, 2007.

[3] Buckingham, F. Crockett, K. Bandar, Z. O'Shea, J. MacQueen, K. and Chen, M. Measuring Human Comprehension from Nonverbal Behaviour using Artificial Neural Networks, Proceedings, WCCI 2012 IEEE World Congress on Computational Intelligence Australia, pp368-375, 2012 ISBN 978-1-4673-1489-3, DOI: 10.1109/IJCNN.2012.6252414

[4] Dolara, A., Grimaccia, F., Leva, S., Mussetta, M., Ogliari, E. A Physical Hybrid Artificial Neural Network for Short Term Forecasting of PV Plant Power Output, Energies, Vol (08), Pp. 1138-1153, 2015, DOI:10.3390/en8021138.

[5] Illias, H A., Chai XR, Abu Bakar, A H., Mokhlis H. Transformer Incipient Fault Prediction Using Combined Artificial Neural Network and Various Particle Swarm Optimisation Techniques. Li D, ed. PLoS ONE, Vol 10(06), 2015, DOI:10.1371/journal.pone.0129363

[6] Aichouri, I., Hani, A., Bougherira, N., Djabri, L., Chaffai, H., Lallahem, S. River Flow Model Using Artificial Neural Networks, Energy Procedia, Vol(74), Pp. 1007-1014, 2015, DOI: https://doi.org/10.1016/j.egypro.2015.07.832.

[7] Jerić, S. V., Primorac, M. Data mining for assessing the credit risk of local government units in Croatia, Croatian Operational Research Review CRORR, Vol(8), Pp. 193–205, 2017, DOI: 10.17535/crorr.2017.0012.

[8] Monika., Sangwan, O.P. Software effort estimation using machine learning techniques," *2017 7th International Conference on Cloud Computing, Data Science & Engineering - Confluence*, Noida, Pp. 92-98. 2017, DOI:10.1109/CONFLUENCE.2017.7943130, URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7943130&isnumber=7943112.

[9] Wang, J., Song, J., Chen, M., Yang, Z. Road network extraction: A neural-dynamic framework based on deep learning and a finite state machine. Int. J. Remote Sens, Vol (36), Pp.3144–3169, 2015, DOI: https://doi.org/10.1080/01431161.2015.1054049.

[10] Tahmasebi, P., Hezarkhani, A. A hybrid neural networks-fuzzy logic-genetic algorithm for grade estimation, Computers & Geosciences, Vol(42), Pp. 18-27, 2012, DOI: https://doi.org/10.1016/j.cageo.2012.02.004.

[11] Schneider, Michael, and Peter Edelsbrunner. "Modelling for Prediction vs. Modelling for Understanding: Commentary on Musso et al.(2013)." Frontline Learning Research 1.2, pp99-101, 2013

[12] Quinlan, J.R., C4.5 program for machine learning, Morgan Kaufmann Publishrs, 1993.

[13] Ahmad, M. W., Mourshed, M., Rezgui, Y. Trees vs Neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption, Energy and Buildings, Vol(147), Pp. 77-89, 2017, DOI: https://doi.org/10.1016/j.enbuild.2017.04.038

[14] Curram, S. Mingers, J. Neural Networks, Decision Tree Induction and Discriminant Analysis: An Empirical Comparison, The Journal of the Operational Research Society, Vol. 45, No. 4. , pp. 440-450, 1994

[15] Buckingham, F. Crockett, K. Bandar, Z. O'Shea, J. FATHOM: A Neural Network-based Non-verbal Human Comprehension Detection System for Learning Environments, 2014 IEEE Symposium on Computational Intelligence and Data Mining (CIDM), pp. 403 – 409, DOI: 10.1109/CIDM.2014.7008696, 2014

[16] Manusov, V. and Trees, A. "Are you kidding me?": the role of nonverbal cues in the verbal accounting process". Journal of Communication, vol. 52(3), pp. 640-656, 2002.

[17] Knapp, M. Hall, J. Nonverbal Communication in Human Interaction, 3rd Ed., Fort Worth: Harcourt Brace, 1992.

[18] Ekman, P. Freisen, V. The Facial Action Coding System (FACS), Consulting Psychologists Press, Palo Alto, CA, US, 1978.

[19] N. Dael, M. Mortillaro, and K. R. Scherer, "The body action and posture coding system (BAP): development and reliability". Journal of Nonverbal Behavior, vol. 36(2), pp. 97-121, 2012.

[20] Turabzadeh, S., Meng, H., Swash, R. M., Pleva, M., Juhar, J. Real-time emotional state detection from facial expression on embedded devices, 7th *International Conference on Innovative Computing Technology (INTECH)*, Luton, Pp. 46-51, 2017, DOI:10.1109/INTECH.2017.8102423.

[21] Savran,A., Alyüz, N., Dibeklioğlu, H., Çeliktutan, O., Gökberk, B., Sankur, B., Akarun, L., Bosphorus. Database for 3D Face Analysis, Biometrics and Identity Management, Lecture Notes in Computer Science Volume 5372, pp 47-56, 2008

[22] Yin, L., Chen, X., Sun, Y., Worm, Y., Reale, M., A High-Resolution 3D Dynamic Facial Expression Database (2008), 8th IEEE International Conference on Automatic Face & Gesture Recognition, FG '08.

[23] Bartlett, M.S., Hager, J.C., Ekman, P., and Sejnowski, T.J. Measuring Facial Expressions By Computer Image Analysis, Psychophysiology, 36, pp.253-263, 1999.

[24] Porter, S. Brinke L., Wallace, B , Secrets and Lies: Involuntary Leakage in Deceptive Facial Expressions as a Function of Emotional Intensity, J Nonverbal Behav, 36:pp23–37, 2012, DOI 10.1007/s10919-011-0120-7

[25] Valstar, M. Pantic, M. (Fully Automatic Recognition of the Temporal Phases of Facial Actions IEEE Transactions on systems, man, and cybernetics—part b: cybernetics, vol. 42:2, 2012.

[26] Waring, H. Z. Expressing noncomprehension in a US graduate seminar. Journal of Pragmatics, Vol. 34(12), pp. 1711-1731, 2002.

[27] Kalliatakis, G., Stergiou, A., Vidakis, N. Conceiving Human Interaction by Visualising Depth Data of Head Pose Changes and Emotion Recognition via Facial Expressions, Computers, Vol. 06(25), pp. 1-16, 2017, DOI:10.3390/computers6030025, pp 1-16.

[28] Brock,S.,Knapp,P. Reading comprehensionabilitiesofchildrenwithattention-deficit/hyperactivitydisorder. Journal of Attention Disorders,1, pp.173–186, 2006

[29] Kintsch, W. Comprehension: A paradigm for cognition. New York: Cambridge University Press, 1998

[30] McNamara, D. S., Magliano, J. P. Towards a comprehensive model of comprehension. In B. Ross (Ed.). The psychology of learning and motivation, 51, Elsevier Science, pp. 297-284, 2009.

[31] Coniam, D. "The use of audio or video comprehension as an assessment instrument in the certification of English language teachers: A case study." System 29.1, pp.1-14, 2001.

[32] Rossi, Michael J., et al. Video informed consent improves knee arthroscopy patient comprehension. Arthroscopy: The Journal of Arthroscopic & Related Surgery 21.6 , pp. 739-743, 2005.

[33] Bowdle, B. F., & Gentner, D. The career of metaphor. Psychological Review, 112, pp. 193–216, 2005

[34] Van Mulken, Margot, Rob Le Pair, and Charles Forceville. The impact of perceived complexity, deviation and comprehension on the appreciation of visual metaphor in advertising across three European countries. Journal of Pragmatics 42.12 , pp. 3418-3430, 2010

[35] Family Health International 360, 2015, Who We Are. [Online] Available: https://www.fhi360.org/. [Accessed 31/1/2018]

[36] National Institute for Medical Research Tanzania, 2011, About Us [Online] Available: http://www.nimr.or.tz/. [Accessed 31/1/2018]