e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:07/Issue:01/January-2025 Impact Factor- 8.187 www.irjmets.com

PERSON BIOMETRIC IDENTIFICATION THROUGH EEG SIGNALS

Maaz Alam*1, Rabia Waqar*2

*1National Center For Big Data And Cloud Computing (NCBC), UET Peshawar, KP, Pakistan.

*2Department Of Computer Systems Engineering, UET Peshawar, KP, Pakistan.

DOI: https://www.doi.org/10.56726/IRJMETS65847

ABSTRACT

EEG-based biometric identification has emerged as a highly reliable and non-invasive modality for individual recognition through the evaluation of unique patterns of brain activity. While promising, getting solid performance and high accuracy is still very difficult. So, this research aims at the feature extraction improvement and utilizes Siamese convolutional 1D neural networks to detect individual differences based on frequency patterns in EEG data. After rigorous investigation, two experimental configurations were identified: The model was trained on a single task and tested on multiple tasks with the ability to generalize across different brain activities. The following analysis showed that the recognition accuracy was above 99%, which emerged the effectiveness of the proposed method. This study improves upon the existing work on EEG biometric systems by presenting one that is both highly efficient and secure; hence, it demonstrates the power of deep learning for creating novel identification systems.

Keywords: EEG Biometric, Siamese Model, Delta Waves, 1D Neural Networks.

I. INTRODUCTION

In other words, a signal is any data that is transmitted or received or processed. Signals are generally electrical or electromagnetic waves in electronics and communication, that originate from various sources, such as audio, video, radio, and digital signals. Another type form is the electroencephalogram (EEG). EEG signals have become a key area of research in modern neuroscience because they shed light on many cognitive and physiological mechanisms. These signals must be measured and interpreted in order to learn how the brain works and identify neurological conditions, signals created by the brain via its electrical activity. EEG signals are research in many different processes of the brain as well as in human sleep, remembering, attention or any other brain operation. The more insight we have on all these neurological principles at work and behind all these brain functions, scientists and re searchers can develop different methodologies on how to treat various different neurological disorders and find alternative measures that are more effective to enhance cognitive performance. The electrical activities of human brain are recorded using EEG signals by placing-down electrodes on the head. Because they reveal activity patterns in different areas of the brain, the resulting signals can be used to study a range of brain functions and acts. Due to their non-invasiveness and the large amounts of information they can provide about the brain, the importance of EEG signals has increased significantly in recent years. Along with its application to brain research, EEG signals are also explored for various biological applications such as biometric identification. Biometric identification systems identify individuals based on their unique physical or behavioral characteristics. Face recognition, iris scanning, and fingerprinting are the most commonly used biometric identification methods. These traditional biometric identification methods have their limitations; possibility of false positive or fake identification, need of physical contact, inability to recognize, if one is wearing a face mask or other face coverings etc. On the other hand, EEG waves as biometrics identification also has some benefits. For biometric identification, EEG signatures are an ideal because, as a personal attribute of each human, they can be easily captured non-invasively. Despite implementing the limitations in traditional biometric identification methods, EEG-based biometric systems offer reliable identification for authentication. Machine learning techniques are 1Frequently applied for classifying EEG data in biometric applications. One such method is the Siamese model, a neural network architecture which has been shown to be effective in several classification tasks. The Siamese model is very helpful for comparing pairs of EEG data to determine if reaching them comes from the same person or not. This study aims to explore the Siamese model to be able to classify EEG waveforms for biometric identification. The model is evaluated on a dataset of EEG signals obtained from a group of participants to test its effectiveness and is compared with other

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:07/Issue:01/January-2025 Impact Factor- 8.187 www.irjmets.com

classification models. The study will assist how an EEG based biometric Identification system can be used in a further reliable way.

II. METHODOLOGY

In this section, we will talk in-depth about the proposed methodology. The first stage of proposed methodology will be from dataset selection. Introduced various steps of cleaning the dataset and getting it ready for the next operations to be done. Now, we will talk about the siamese architecture that we use for our research work.

Dataset

In this study, we employed the EEG Motor Movement/Imagery Dataset, which is a publicly available and reputable dataset made available by the PhysioNet repository. The dataset includes more than 1,500 EEG records, each lasting from one to two minutes, making it a large and heterogeneous dataset for analysis. The wide variety of subjects (109 total) also allows for real-world deep learning model generation - engineered to generalize and transfer well to new scenarios. Data comes from various subjects, covering a substantial amount of motor movements and mental imagery patterns, making it a rich dataset suitable for feature extraction. This holistic strategy enables the development of deep learning algorithms that can accurately model the complex features of brain function experienced during these tasks. Recordings were performed with the BCI 2000 system using 64-channel EEG data collection. This produced a diverse and rich data set, with each participant running 14 experimental runs. The EEG data were collected according to the 10-10 system of electrode placement with certain exclusions, presented in the associated figures and tables. For this research, we only used the baseline “eyes closed” portion of the recordings, as it was simple to isolate and replicated widely, in addition to being readily usable in biometric identification. This innovative segment uniquely encapsulates the individual brain activity and enables a reliable and replicable approach for identifying individuals.

Preprocessing

Dataset preparation, the processing of EEG signals, is a significant step, since EEG signals generated by the brain also include signals recorded from more than one source. Even low-level eye movements, or physiological activities like heart beat or muscle movements can generate significant noise. Such interferences must be addressed in order to train relevant AI models. One noteworthy study is conducted by N. Bigdely et al. There is a very powerful large-scale EEG preprocessing pipeline [12] capable of obtaining high results while keeping valuable information of the signal intact. There are four steps in the pipeline:

1. Line Noise Removal — techniques for removing electrical interference (line noise) from EEGs This is an essential step to ensure the fidelity of the data and remove artifacts that may interfere with analysis.
2. With respect to strong references based on true average signals, true average reference signals are adopted to remove much of the common noise source to offer increased quality and define brain activity.
3. Noise-filtered channels referenced through a comparative analysis alert for abnormal or inconsistent activity, thus securing channels with only reliable references for additional processing.
4. Maintain Data Integrity for Cross-Referencing — To ensure the integrity of each individual user's dataset for transparent analysis, a sufficient amount of data is saved for all users to allow for reverse interpolation techniques to be used. In this case, we save all data for at least the last 15 users.

This preprocessing approach ensures the dataset is already preprocessed and ready for tasks like feature extraction and AI model training. By mitigating noise, refining signals, and preserving essential data, the pipeline creates a high-quality dataset suitable for meaningful EEG research.

Visual Impact of Preprocessing:

Figure [2] compares the raw EEG signal and the processed EEG signal. The “before” image reveals interference from noise sources like line noise, muscle artifacts and eye movements which mask the patterns of brain activity. With preprocessing applied, this image illustrates the drastic improvements, which lead to a much cleaner signal and lower artifact presence. Overall, this was also a clear improvement demonstrating that this pipeline does help us to obtain higher quality EEG data, thus leading to better, more reliable science.

Frequency band extraction

Once we have pre-processed the dataset using the prep pipeline, we want to get frequency bands from the EEG signals. The electroencephalography (EEG) signals are exampled as five major frequency bands representing

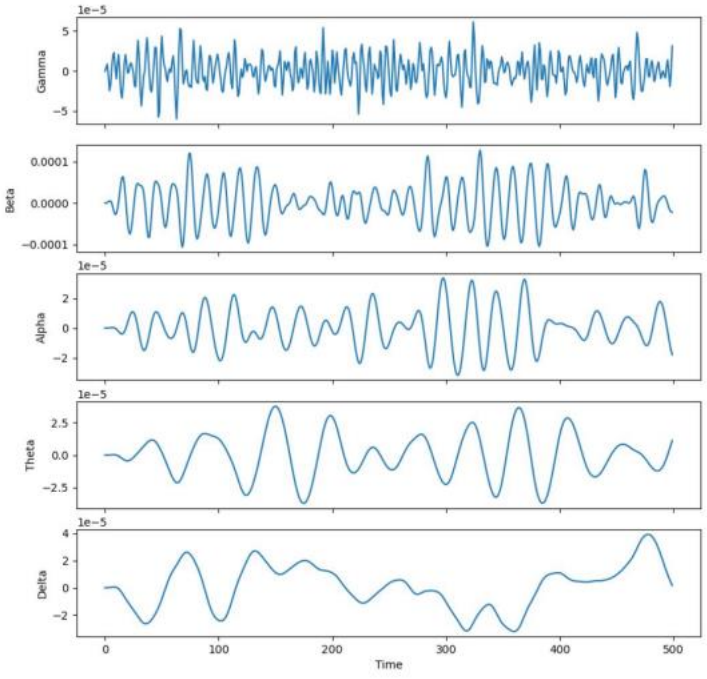
e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:07/Issue:01/January-2025 Impact Factor- 8.187 www.irjmets.com

significant brain activities. These bands– Alpha, Beta, Gamma, Delta, and Theta– are essential, since nearly all brain functions linked with its ranges. These bands are displayed in Fig [3], while their frequency ranges are summarized in Table [2]. Chapter 4: Neuroanatomy explains the significance of those bands: Each band is connected to specific physiological and cognitive functions:

1. Delta Waves (0.5–4 Hz): The slowest wave, with the greatest amplitude. They are most common in deep sleep and in some neurological disorders. Delta waves are associated with deep relaxation, healing, and growth hormone release and are essential to working of restorative functions in brain.
2. Theta Waves (4-8 Hz) — These waves occur in light sleep and drowsiness and during meditation, and in children and people with attention deficit disorders. They're related to creativity, intuition, and deep emotional states.
3. Alpha Waves (8-12 hertz): Waves are a well-studied frequency range associated with relaxed states like daydreaming and meditation. They represent relaxed wakefulness and often happen at peaceful and meditative times.
4. Beta Waves (12-30 Hz): Low-amplitude wave associated with alertness, concentration, and active mental states. These are associated with concentration and alertness as well as problem-solving abilities.
5. Gamma Waves (30–100 Hz): Gamma waves are quickest and are really observed during cognitive processing, sensory perception. These rhythms are crucial for higher mental functions like learning, perception and awareness.

These frequency bands are instrumental in the analysis of brain activity and its associated relations to different mental and physiological states.

EEG Bands	Frequency Range
Delta	< 4Hz
Theta	4 – 8Hz
Alpha	8 – 13Hz
Beta	13 – 30Hz
Gamma	> 30Hz



Significance of Delta Pattern

Methods utilizing Delta in EEG-based biometric systems are not widely explored, and this great potential for subject differentiation requires thorough investigation. Traditionally, research has been centered on Alpha and

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:07/Issue:01/January-2025 Impact Factor- 8.187 www.irjmets.com

Beta patterns, as they occur when someone is relaxed and focused. These bands were found to provide the most unique data for biometric identification. However, recent work from Xiang et al., 2018) has pointed out the model to show promise for classification and subject identification with the Delta pattern.

Historically, Alpha and Beta patterns have been instrumental in investigations seeking to differentiate changes in brain function between resting and focused periods. Capitalizing on these shifts, researchers created systems to tell people apart using these bands.

Qualitative Analysis:

Given that the signals were different for the subjects, a pairwise analysis based on cosine similarity and correlation metrics between the brain signals from different subjects was carried out to find the distinguishing feature of EEG signals for user identification.

Cosine Similarity:

Six cosine similarity measures were calculated, as shown in Figure [4]. Among these, the Delta pattern exhibited the lowest cosine similarity. This indicates its distinctiveness for biometric identification. It suggests that Delta patterns offer unique features ideal for accurate and reliable user differentiation.

Correlation Analysis:

An inter-subject correlation analysis was performed to identify the most effective pattern for distinguishing between subjects. The process involved:

1. Pick 8 random subjects and collect 10 samples from the first subject.
2. Gathering 10 samples per remaining 7 subjects (i.e. having 70 samples total).
3. Computing 700 pairwise correlations between the samples from the first subject and those from all other subjects.
4. Calculating an inter-subject correlation coefficient for the first subject by taking the average across these correlations.
5. Doing this for all frequency patterns and check how effective they are at differentiating.

Siamese Model

The Siamese model is a neural network architecture which incorporates a pair of inputs and was developed by Bromley, Guyon, LeCun, Säckinger and Shah in the 1990s. Siamese network are the names given to its "twin" sub-networks that share the same parameters and weights; it was first designed for signature verification, to classify whether two signatures were from the same person.

The architecture input paired signature images to the sub-networks and their outputs were evaluated using a distance metric for signature classification. Siamese models became popular for face recognition systems in the early 2000s, which worked by looking at two images of faces and determining whether they were from the same person or not. Its uses in the image domain continued evolving, with applications in image retrieval, text similarity assessment, and more recently, NLP tasks such as question-answering and paraphrase detection.

Deep learning developments further generalized the Siamese architecture by employing convolutional neural networks and recurrent neural networks to further expand the range of real-world use cases. Now Siamese networks are being used by industries like finance, healthcare, e-commerce, etc. for tasks like fraud detection, recommendation systems, etc. Similarly, Siamese networks study text similarity by encoding the semantic meanings of sentences or paragraphs into a vector representation of a fixed size. Using a metric such as cosine similarity, the degree of similarity between vectors can be calculated, allowing the accurate identification of their relationship with one another on the basis of the content.

Siamese Architecture in This Study

Figure [6] illustrates the Siamese net used in this study trained to compare input pairs for determining similarity and dissimilarity. An example of such architecture will be to have two identical base networks with shared weights where they learn similar features and is followed by a distance metric that measures the similarity Each of the base networks receives input sequences of shape (64, 1) and then passes them through a sequence of layers until a fixed-length feature vector is formed.

Processing Workflow

a. Features Vectors: The base networks are used to get feature vectors for the left and right inputs.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:07/Issue:01/January-2025 Impact Factor- 8.187 www.irjmets.com

b. A custom function is used to compute the Euclidean distance to compare these vectors.
c. The output of the distance function is a scalar value flattened and passed through dense layer with the sigmoid activation function, returning the similarity of input pair between 0 to 1.
The implementation is done using Python's Keras library, where input, output, loss function, optimizer, and evaluation parameters are defined in functions.

Base Network Architecture

The base network summarized in Figure [7] takes a sequence of 64 elements and outputs feature vectors. It is a multi-layer convolutional structure with ReLU activations and fully connected layers. Objective of the component is to identify most informative features of input sequences used to calculate similarity.

A sigmoid activation then scales the Euclidean distance between the feature vectors of two input sequences (left and right branches of the network) to a value in the interval [0,1] to evaluate similarity.

The network is trained by minimizing a loss function, which measures difference between predicted similarity and actual similarity.

Mathematical Explanation of the model

Let xxx and yyy represent two input sequences, and f(x)f(x)f(x) and f(y)f(y)f(y) denote their corresponding feature vectors generated by the base network.

The base network composed of multiple convolution and fully connected layers, defined as follows:

$h_1(x) = \text{ReLU}(\text{conv}_1(x))$

$h_2(x) = \text{ReLU}(\text{conv}_2(h_1(x)))$

$h_3(x) = \text{ReLU}(\text{conv}_3(h_2(x)))$

$h_4(x) = \text{ReLU}(\text{conv}_4(h_3(x)))$

$h_5(x) = \text{FC}_1(h_4(x))$

$h_6(x) = \text{ReLU}(\text{FC}_2(h_5(x)))$

$h_7(x) = \text{ReLU}(\text{FC}_3(h_6(x)))$

$h_6(x) = \text{FC}_4(h_7(x))$

where conv_i denotes the i-th convolutional layer, FC_i denotes the i-th fully connected layers, and ReLU is rectified linear unit activation functions. Compute the similarity between the two feature vectors, the Euclidean distance is used

$$D(x, y) = s\_\sum_{i=1}^{n} f(y)_i)^2 \\ (f(x)_i -$$

In this equation, n is the dimension of the output of the base network (i.e., number of units in the last layer), f(x) and f(y) are the output vectors of the base network for inputs x and y, respectively, and the sum is taken over all i dimensions of the output vectors. Finally, the output of the network is obtained by applying a sigmoid activation function to the distance:

$\text{output} = \sigma(D(x, y))$

Where σ is the sigmoid activation function and D(x, y) is the Euclidean distance between the output vectors of the base network for two input samples x and y. During training, the parameters in the network are adjusted to minimize the difference between output and the true similarity S(x,y) between the two input sequences. This is typically done by minimizing a loss function L:

$L = (S(x, y) - \text{output})2$

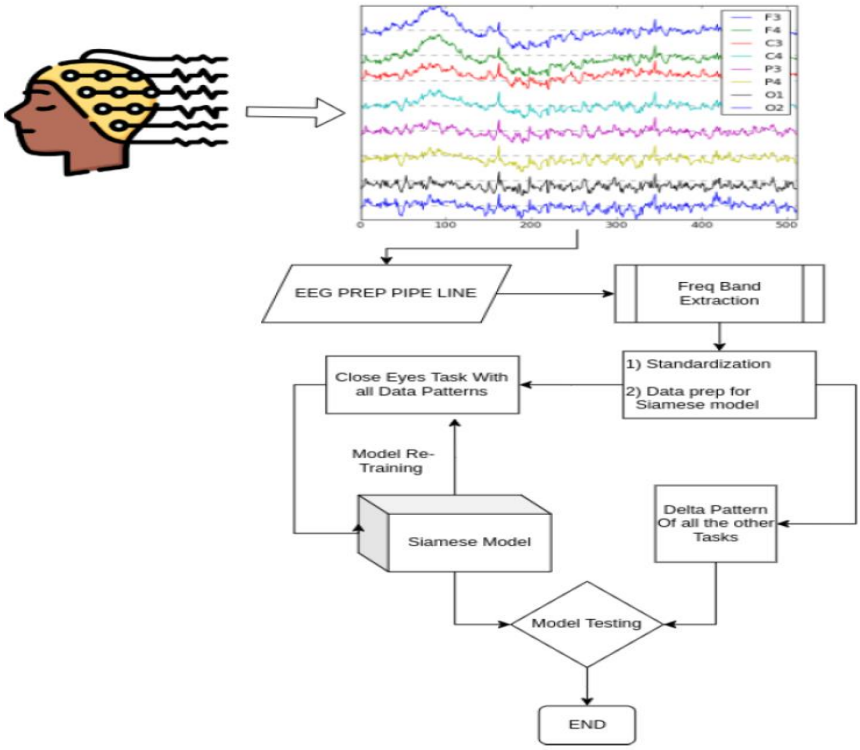
In this equation, S(x,y) is the similarity score between input samples x and y, which is computed as the negative Euclidean distance between their output vectors from the base network. The output of the Siamese network is then obtained by passing this similarity score through a sigmoid activation function. The goal of the loss function is to minimize the difference between the predicted output and the actual similarity score.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:07/Issue:01/January-2025 Impact Factor- 8.187 www.irjmets.com

III. MODELING AND ANALYSIS

Model Hyper parameters

Model used binary cross entropy loss with Adam optimizer for classifying input samples into two classes. Model is trained for twenty epoch, with a batch size of 32 and 10% of the data is used for validation. Reduce LR Plateau callback (implemented in Keras) intelligently decreases the learning rate if there is a plateau in the accuracy of the selected metric to promote faster convergence. Also, Early Stopping callback is used to stop our model from overfitting by stopping the training if our validation metric (i.e, val accuracy) is not improved for a specific number of epochs. Tweaking these hyper params becomes really important to improve model performance and generalization. The Motor Imagery dataset was originally acquired with the Python library MNE, an open-source toolbox enabling human brain signals analysis. An optimal setup was reached MNE to read, process, and visualize neurophysiological data (e.g., electroencephalograms or EEG configurations) and other formats such as. edf files. The model was validated by 0.1% of from the EEG prep pipeline, which is explained in Proposed Methodology. Subsequently, the advanced data was segregated to produce a total of five different types of EEG training data, which was retained as the validation set. Mech and MEG are used with the same model architecture and hyper. Dataset was preprocessed after retrieval frequency bands Training and testing datasets were formed for each band, based on data from 70 subjects for training and 39 subjects for testing. The Siamese model was trained with various hyper parameter configuration were applied consistently across all six experimental setups.



IV. EXPERIMENTS AND RESULTS

Experimental Setups

Evaluation was performed on two experimental setups. The first setup was training and testing specifically on the relaxation eye closure detection task. This setup provided baseline performance when subjects had their eyes closed. In the second setup, we evaluated the model's generalization capabilities by recognizing and classifying subjects engaged in different activities from the mere closure of the eyes. This wide-ranging assessment was designed to measure model's diversity and flexibility in practical situations. Combined, these setups offered a comprehensive view of the model performance in both controlled and real-world conditions, emphasizing its effectiveness and potential use case.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:07/Issue:01/January-2025 Impact Factor- 8.187 www.irjmets.com

Evaluation Metrics

Several model evaluation metrics accuracies including, recall, precision and F1 score are used to measure the performance of the proposed model. Together this gives us a complete overview regarding model ability to correctly identify and predict the target variables. Below is a brief description of each metric:

Accuracy:

Accuracy measure overall correctness of the model prediction. Its the ratio of the correctly classified instance to the total instances in dataset. Mathematically:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

Here, TPTP (true positives) are the instances correctly classified positive, TNTN (true negatives) instances correctly classified as negative, FPFP (false positives) are instances incorrectly classified positive, and FNFN (false negatives) are instances incorrectly classified as negative.

Recall:

Also referred to as sensitivity or true positive rate, recall measures model ability to correctly identify the positive instances. It is calculated as:

$$\text{Recall} = \frac{TP}{TP+FN}$$

Recall indicates the proportion of actual positive instances correctly identified by the model.

Precision:

Precision evaluates the model's accuracy in predicting positive instances. It measures the proportion of true positive instances among all instances predicted as positive:

$$\text{Precision} = \frac{TP}{TP+FP}$$

F1 Score:

F1-score is harmonic mean of precision and recall. It offers balanced measure of the model performance. It is particularly useful when there is an uneven class distribution. The formula is:

$$\text{F1 Score} = 2 \cdot \frac{Precision \cdot Recall}{Precision+Recall}$$

The F1 score accounts for both precision and recall. This ensures a balance between correctly identifying positive instances and decreasing false positives.

By implementing these metrics, we can check our model's performance. The insights derived from accuracy, recall, precision. F1 score helps us evaluate the model's effectiveness in classification and prediction tasks. These results guide our analysis and support meaningful conclusions based on experimental findings.

V. CONCLUSION

The research mentioned focuses on using EEG signals for motor imagery-based BCI. The initial step of this methodology consisted of acquiring the Motor Imagery dataset through the MNE Python package, which is an advanced open-source tool to preprocess and analyze brain signals, utilizing the EEG prep pipeline. A Siamese neural network model was used in the study to classify EEG signals to identify the difference between such inputs. After tuning multiple hyperparameters, we were able to arrive at an optimal configuration. Its performance was validated using 0.1% validation data hidden from training set. The study is structured around two different experimental setups to evaluate the performance and generalizability of the model. In Setup 1, the model was trained and tested on just one class (closed-eye relaxation), which was used as a baseline, yielding a 99% accuracy on the task. Setup 2 was the test of adaptability and robustness of the model to classify subjects performing other activities than just eyes closed relaxation. The results showed a strong degree of generalization of the model, which achieved high accuracy on tasks, even those, of which there are very few examples in the dataset. And these findings prove that, in terms of identifying the patterns on the EEG eigen-spectra of subjects to differentiate them irrespective of the tasks carried out (i.e. whether at different tasks or

not), the Siamese 1D ConvNN architecture is effective in learning valid high level/abstract information from the tasks. These findings demonstrate the value of exploring delta frequency patterns and Siamese Conv1D models for EEG-based biometric identification. This accuracy in closed-eye prediction in addition to good generalization across tasks could potentially translate to true positive rates (TPR) for robust and secure authentication in the wild. Simultaneously, the data-driven model was integrated into a RL-based human-centric energy management system. The system displayed intelligence, adaptability and energy efficiency while also taking advantage of the novelty of other parameters to match occupant needs, and stressed sustainability through energy conservation. This section outlined the experimental setup, evaluation metrics, parameter configuration, results, and evaluation of the proposed system's performance.

VI. REFERENCES

[1] T. Mohana Priya, Dr. M. Punithavalli & Dr. R. Rajesh Kanna, Machine Learning Algorithm for Development of Enhanced Support Vector Machine Technique to Predict Stress, Global Journal of Computer Science and Technology: C Software & Data Engineering, Volume 20, Issue 2, No. 2020, pp 12-20

[2] Ganesh Kumar and P.Vasanth Sena, “Novel Artificial Neural Networks and Logistic Approach for Detecting Credit Card Deceit,” International Journal of Computer Science and Network Security, Vol. 15, issue 9, Sep. 2015, pp. 222-234

[3] Gyusoo Kim and Seulgi Lee, “2014 Payment Research”, Bank of Korea, Vol. 2015, No. 1, Jan. 2015.

[4] Chengwei Liu, Yixiang Chan, Syed Hasnain Alam Kazmi, Hao Fu, “Financial Fraud Detection Model: Based on Random Forest,” International Journal of Economics and Finance, Vol. 7, Issue. 7, pp. 178-188, 2015.

[5] Hitesh D. Bambhava, Prof. Jayeshkumar Pitroda, Prof. Jaydev J. Bhavsar (2013), “A Comparative Study on Bamboo Scaffolding And Metal Scaffolding in Construction Industry Using Statistical Methods”, International Journal of Engineering Trends and Technology (IJETT) – Volume 4, Issue 6, June 2013, Pg.2330-2337.

[6] P. Ganesh Prabhu, D. Ambika, “Study on Behaviour of Workers in Construction Industry to Improve Production Efficiency”, International Journal of Civil, Structural, Environmental and Infrastructure Engineering Research and Development (IJCSEIERD), Vol. 3, Issue 1, Mar 2013, 59-66