Indonesian Journal of Electrical Engineering and Computer Science
Vol. 33, No. 1, January 2024, pp. 634~642
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v33.i1.pp634-642

Human addictive behavior prediction by using lime with ensemble model

V Sabapathi[1], Selvin Paul Peter Jacob[2], Woothukadu Thirumaran Chembian[3], Kandasamy Thinakaran[4]

[1]Department of Computer Science and Engineering, Vel Tech High Tech Dr. Rangarajan Dr. Sakunthala Engineering College, Chennai, India
[2]Department of Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, India
[3]Department of Computer Science and Engineering, Vel Tech High Tech Dr. Rangarajan Dr. Sakunthala Engineering College, Chennai, India
[4]Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, India

Article Info

Article history:

Received Aug 9, 2023
Revised Nov 14, 2023
Accepted Nov 21, 2023

Keywords:

Decision making
Drug addict
Ensemble classifier
Human behavior pattern
LIME

ABSTRACT

The data-driven techniques have utilized data mining and machine learning (ML) techniques in the biomedical and healthcare fields. The process of decision-making in uncertain contextual related to human addictions and emotions play an important role in the present research. The main aim of the research is to perform classification and generate a support system for uncertain addiction circumstances by proposing a technique for drug addiction treatment. The human behavior has majority shown challenges for the prediction of human behaviors that includes body poses estimation, movements and interaction with objects. This pose estimation has showed complexity with more pose aspects and the proposed research attempts to understand the human behaviors. The present research uses the local interpretable model-agnostic explanations (LIME) for finding the input features which are most important to generate a particular output based on decision service. LIME understands the model to perturb the data samples as an input and understands shows predictions change. Also, the ensemble classifier contains classifiers group that combines for performing the prediction of all unseen instances based on voting. The proposed LIME Feature-Ensemble classifier obtained 97.54% of accuracy when compared to the existing convolutional neural network (CNN) of 59.33% and Ensemble model of 93.33% accuracy.

Corresponding Author:

V Sabapathi
Department of Computer Science and Engineering, Vel Tech High Tech Dr. Rangarajan Dr. Sakunthala Engineering College
Chennai, India
Email: sabapathi@velhightech.com

1. INTRODUCTION

The drug is a chemical or medical constituents that modifies the brain functions resulting with mood perception changes, consciousness, behaviors, and cognition [1]. The drug causes serious harm among people brains and bodies [2]. Therefore, a criminal act has caused because of these endangering qualities that has involved statistics and probability theory, complexity in the algorithms working with other disciplines [3]–[5]. The model has extracted useful information which are hidden incompletely due to random noise generation. The existing data mining techniques use an iterative process which showed complexity based on the continuous cyclic process [6], [7]. The decision-making based on the addictive behaviors is ambiguous which is necessitating self-awareness [8]. However, there is still more research required to investigate the needs of the human psychometric model which quantifies the activities of a particular area of addiction [9]–[11].

The addiction is transformed from one person's physiological characteristics and addictive behavior is compulsive behavior. It begins as a pleasurable activity that progresses to serious demand and compulsive conduct [12]–[14]. The main goal was to show a framework that allowed an individual to input the lifestyle activities as input [15]–[17]. Sometimes, human behavior is sensitive, compulsive, and also time-consuming [18]. Therefore, it is essential to develop a cognition model for showing self-awareness of the addictive context. As there is unconsciousness present in the occasional activity, the model would show unnoticeable changes impacting the addictive perception [19]. Therefore, the addictive classification to structure various addict contexts using thresholding value, when the addict context is showing difficulty [20]. The quantity parameter is adopted as it is memorable for self-analyzing the quantity in multi contexts. The substance-based additions measure the quantity of consumption that seeks the substance for measuring the quantity of time spent and relevant period [21]. However, the prospect of emerging regulatory skills is further hampered by the leakage into the transient withdrawal of addictive behaviour [22]. This is because it encourages theexploration for instant satisfaction and elevates the behaviour to a chief role in person's life throughdeterminedviews, which helps to maintain the addiction and, ultimately, impedes treatment [23]. Addiction is a multifaceted illness that impairs both brain and bodily functions. Research has taken a variety of approaches to the various psychological components that underlie addiction, leading to the development of interpretative theories and models [24].

The existing researches for behavior addiction detection using machine learning models are as follows: Walters *et al.* [25] used machine learning models for identifying the imminent drinking habit. The prediction model tailored messages for those drinkers who are under high risk of homelessness. The predictors include mood, which were not interacting with the one who is in depression, lower commitment, being alcohol, urge to drink alcohol, alcohol-free that are easily available confident and being indoor. The models do not suggest only EMA reports but were worthy to assess the strong results. Hassanpour *et al.* [26] utilized a deep neural network that identified a substance based on social media data. Deep convolutional neural network (CNN) with long short-term memory (LSTM) was used to extract the predictive features. There are 228 individual 228 features which are related to alcohol addiction that were used for evaluation. The model classified the substance based on the risk but failed to provide explicit insight that has specific elements for increasing the risk of substances based on Instagram posts. Tassone *et al.* [27] utilized graph mining and deep learning models to determine the usage of the drug based on Twitter data. The classification using multiple methods showed better results compared to the existing classifiers. The present research utilized the drug which corresponded to the substance levels that frequently provide the usefulness of the system. Shahriar *et al.* [28] developed a machine learning approach for predicting the vulnerability of drug addiction. The number of the dataset used was larger and appreciated when the locations or areas with data collection were enlarged based on the variations in the outcome.

Lu *et al.* [29] used boosting-based machine learning (ML) algorithms for predicting the level of aggression among the people addicted to drugs. The gradient boosting regression found the relevant psychological variables showed security, capital, trust, and alexithymia that were related to aggressive behavior significantly. However, the model was difficult to establish the true cause and relationships between the aggression and variables. However, the longitudinal data and ML method combined data should be applied. Yuan *et al.* [30] utilized a ML model for treating drug addicts based on virtual reality therapy. The drug addicts were watching virtual reality (VR) videos for distinct scenes, and heart rate and physiological data were collected. The hybrid ML technique combined K-means clustering with principal component analysis (PCA) for determining the drug addicts that find the relationship among the characteristic data. However, physiological characteristics were added and considered for comprehensive evaluation from multiple angles. Local interpretable model-agnostic explanations (LIME) look at what happens to the predictions when the machine learning model is fed different versions of your data. LIME delivers local model interpretability by varying the feature rate for data sample and monitoring the result influence. This often has to do with the queries that individuals have when viewing the output of a model. Ensemble methods are quite useful once a dataset has linear and non-linear data since multiple models are used to handle. For the aforementioned reasons, this study suggested the LIME Feature-ensemble classifier for the treatment of drug addiction. The contribution of the research work is to examine and model the addicted substances by humans using a cognitive psychometric model for decision making. The proposed research work contributes the following process:

- This research collected some of the data sources such as universities, rehabilitation facilities, and addiction treatment facilities which are used to produce the primary quantitative and qualitative results.

- Once the data is gathered then, the LIME method is developed to identify the input features that are most crucial for producing a certain output based on decision service. Also, it comprehends how the model perturbs the input data samples and changes the predictions.
- Additionally, the ensemble classifier trains a collection of classifiers that work together to perform the prediction as all unseen instances.

The structure of the present research work are as follows: section 2 illustrates the process of proposed method. Section 3 represents feature extraction using LIME. Section 4 illustrates about the outcomes and comparisons. Section 5 defines the conclusion of the research.

2. PROPOSED METHOD

The major aim is to develop the prediction model to predict the early addictive habit that overcomes the problem of ambiguity during the classification approach. Figure 1 shows the proposed method block diagram. The data is gathered from every source that includes universities, rehabilitation centers, and addiction treatment centers which are used for the primary data quantitative and qualitative data. There exists a second dataset that is not sufficient for the prediction model.

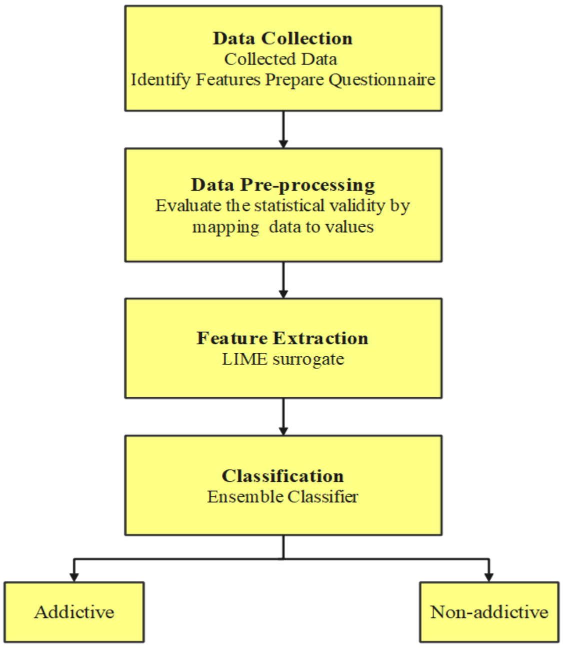

Figure 1. Overview of proposed model

2.1. Data collection

The research work purpose is to find the drug addiction by developing a model for prediction. The gathered data is possible with sources which includes addiction treatment centers, rehabilitation center, and universities. The use qualitative and quantitative primary data methods. They are existed secondary datasets are not sufficient for developing the prediction model. Thus, the dataset is obtained by the HHS.gov official website that mainly focuses on the alcohol consumption, cocaine, cigarettes, and marijuana and their behaviors detection. The database concentrates mainly on 12 attributes which includes ethnicity, gender, age, country, education, personality measurement that have legal and illegal drugs. Also, the questionnaire answers related primarily to stored strings in the CSV files [31]. This later will be converted to numerical values if the symptom present is labeled as 1 or else 0. Researchers use some methods to explain missing data, which leads to several statistical methods for managing missing information. Missing completely at random (MCAR) refers to instances where data are missing at random, for as when multiple questionnaire forms are randomly assigned. Participants may not have access to some variables if several forms are used since they weren't provided with a particular form. Missing at random (MAR), also known as accessible missing data, is another sort of missing data in which a measure that predicts missingness is present yet the data are not fully missing at random. It is important to recognize that a statistical advantage over chance does not always translate into clinical value. Instead, the relative clinical utility of a concept must be considered in relation to the additional cost and potential cost-benefit to the patient. Predictive performance must be statistically measured, regardless of the particular metric, and when employing cross-validation, this should be done via permutation testing.

3. FEATURE EXTRACTION USING LOCAL INTERPRETABLE MODEL-AGNOSTIC EXPLANATIONS (LIME)

With text, image, and tabular data, LIME performs effectively. Fidelity features provides us with a decent indication of the interpretable model's dependability in elucidating black-box forecasts inside the data's interest zone. The key component of LIME is its ability to operate with a readily interpreted, human-readable description of the input. The result of LIME is a collection of explanations showing how each feature affects a data sample's prediction. This allows the feature modifications that will have the most impact on the prediction to be identified and provides local interpretability. In order to put the local surrogate models into practise, LIME was developed. As an alternative to training the model for prediction, the global surrogate model was employed. The black box model, which makes predictions using input data points, is present in the training data. Understanding the ML model for prediction is the primary goal. A fresh dataset consisting of altered samples that match the black box model for the predictions is generated by the LIME. The dataset trains the LIME with an interpretable model. The proximity of the weights is sampled with the instances implemented on the interpretable model. The interpretable model showed good approximation by using a machine learning model for predictions. The local surrogate model showed interpretability constraint which is expressed as shown in (1) and (2).

$$explanation(x) = argmin \in GL(f, g, \pi x) + \Omega(g)explanation \tag{1}$$

$$explanation = argming \in GL(f, g, \pi x) + \Omega(g) \tag{2}$$

From the (1) and (2), the model has an instance x of the model g that is a linear regression which reduces the loss L. It measures that the prediction is close to the original model f while the complexity is low represented as $\Omega(g)$, G is the explanation for the linear regression models. The term 'πx' is known as proximity measure which is defined as a large neighborhood round with an instance x. The LIME optimizes with a loss and the complexity is determined by selecting a number of features at the maximum range. The number of features required for selection is K which is used as an interpretable model. The lower K value is used for easing the interpret model and the higher K produces the higher fidelity. The training models were used for extracting the K features. The Lasso model can monitor the higher regularization parameter λ that can model without features that get weights. These weight estimates differ from the value of zero. If K features are present in the model, then the desired number of features were selected. The other strategies are forwarded and backward to select the features. The full model contains all the features to model the intercept that tests the feature which would show improvement when added or removed until the features are reached. This is dependent on the type of data which can be a text, image, or table. The text images provide a solution with single words when the superpixels are turned off. The LIME creates a new sample to perturb individually the features.

3.1. Classification using ensemble classifier

In contrast, individual models are less reliable predictors than ensemble classifiers. Ensemble methods are quite useful once a dataset comprises linear and non-linear data. Models that use ensemble techniques are usually neither overfitted nor underfitted, and bias and variance can be reduced. A model ensemble has consistently been less noisy and more dependable. Based on behavioural characteristics, drug addiction is classified into classes that are either addictive or non-addictive. The ensemble classifier is employed in the current study project to carry out the behavioural feature classification. Merging several structures with ensemble learning enhances the results and has produced better prediction performances than using a single model. The main idea is to learn the classifiers and make them vote. The models that are used in the present research are K-nearest neighbor (K-NN), random forest (RF), decision tree (DT), and neural network (NN) classifiers [32].

3.1.1. K-nearest neighbor

The present research work uses KNN classifier for the evaluation of distances among the testing and training samples. These are recognized by the nearest neighbors [33]. The KNN classifiers consider only the nearest neighbors based on training and testing samples. There are distinct measures used for the distance calculation among the testing and training samples. It uses Euclidean distance for calculating the distance using the KNN classifier [34].

3.1.2. Random forest

The RF acts as an ensemble classifier that improves the model accuracy [35]. The constructed tree is individually established based on the bootstrap sample which contains actual information. The decision is made on the basis of the performance executed by the tree. Thus, the best decision comes when the tree is able to perform classification based on the voting of class objects. The selected class receives the best number of votes with object numbers. The RF can use both the boosting and bagging processes and selects the random variable for building up the tree [36].

3.1.3. Decision tree

The DT is a classifier that can partition the feature space [37], [38]. The model can consider K number of observations which has p inputs which are related with the response variable $(y_i, x_{i1}, x_{i2}, \ldots, x_{ij}, \ldots, x_{ip})$ for $i = 1, 2, \ldots, K; j = 1, 2, \ldots, p$. The prediction label y_i is having the input features which are represented as i; $(x_{i1}, x_{i2}, \ldots, x_{ij}, \ldots, x_{ip})$ are called as the input features. The classification of tree splits the input to the categories on the basis of input explanatory variables. Each of the variables are partitioned to sub regions that are having few observations. The next criteria when reached processes to continue with feature space partitioned into Q regions $\{R_1, R_2, \ldots, R_Q\}$. The splitting point as a best variable is found when splited into the possible values.

3.1.4. Neural network

The artificial neural network (ANN) model is utilized a ML model which is applied to perform the non-linear process to improve the performance [39]. The multi-layer neural architecture has performed well on the computation model that has neurons connected with the previous layer. The neural network processes two procedure such as backpropagation and forward process. The signals of the input are processed into a forward direction which can be activated with the network layers from input to an output signal. There are 3 layers in FFN model where the output obtained is on the basis of error value that is summed to get an output. The input and the output layers from all the classifiers are combined for enhancing the prediction results based on the voting approach [40].

4. RESULTS AND DISCUSSION

The proposed ensemble model has simulated the results using a MATLAB (2019a) environment. Following that, the system needs to be operated with the windows 10 OS, 8GB RAM, Intel i5 processor, and 4 TB hard disk. Table 1 displays the parameter description.

Table 1. Parameter description

Parameters	Values
Number of attributes	12
Number of split	3
Number of trees	30
Iteration count	100
Desired members	15
Subcommittees count	3
Number of instances	690

4.1. Performance metrics

The proposed LIME feature-ensemble classifier is used in human additive behavior prediction to enhance the overall performance metrics. The mathematical expressions of accuracy, MCC, F-score, specificity and sensitivity are represented in (3) to (8). All the formulas are calculated based on true positive, true negative, false positive and false negative.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \quad (3)$$

$$F - score = \frac{2TP}{2TP+FP+FN} \times 100 \quad (4)$$

$$Specificity = \frac{TN}{TN+FP} \times 100 \quad (5)$$

$$Sensitivity = \frac{TP}{TP+FN} \times 100 \quad (6)$$

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

$$Recall = \frac{TP}{TP+FN} \quad (8)$$

where, TP denotes true positive, FN refers false negative, TN denotes true negative, and FP refers false positive.

4.2. Quantitative analysis

The results obtained by the proposed ensemble model with a local surrogate have obtained better values of accuracy compared to the existing models. The existing models such as NN, support vector machine (SVM), RF, and KNN model obtained distinct performances. The SVM model obtained 92% of accuracy, sensitivity of 94%, specificity of 90.67 %, F-score of 92%, precision of 92%. The results obtained from the proposed methods are shown in Tables 2 and 3. The KNN classifier obtaining 90.24% of accuracy, a sensitivity of 89.70%, and specificity of 88.78%, F-score of 87.25%, and precision of 89.45%. The ensemble model improved the classification which improved the prediction accuracy.

Table 2. Outcomes of proposed model in terms of accuracy, sensitivity, and specificity

Algorithm	Accuracy (%)	Sensitivity (%)	Specificity (%)
Proposed ensemble	97.54	96.78	96.73
NN	92.72	94.12	91.25
SVM	92.0	94	90.67
RF	94.00	97.02	91.38
KNN	90.24	89.70	88.78

Table 3. Results obtained by the proposed method in terms of F-score and precision

Algorithm	F-score (%)	Precision (%)
Proposed ensemble	96.05	97.05
NN	93.00	93.00
SVM	92	92
RF	94.00	92.00
KNN	87.25	89.45

4.3. Comparative analysis

Table 4 is the comparative analysis among existing and proposed method and results obtained by the existing models such as LSTM, CNN, ensemble model, and ensemble models. The value of precision for the LSTM model was obtained as 68.6%, 95.4% for the CNN model, ensemble model at 93%. The existing LSTM model failed to explicit a few of the specific elements during the classification of substances that obtained lower precision of 68.6%. The CNN model showed corresponding and frequently used illicit substances but lacked in training the data that obtained 59.33% of accuracy. Similarly, the dataset is expected to have larger data when the areas of data collection have improved. The results with variation in the results compared with the ensemble learning model showed 97.54% of accuracy.

4.4. Discussion

The study's main objective is to develop a prediction model that can detect early addicted behaviours and resolve the ambiguity problem with the classification approach. Working with an interpretable description of the input that is simple for humans to understand which is a crucial significance for LIME. Since multiple models are connected to handle all data, ensemble methods are quite useful. Individual models are less reliable predictors than ensemble classifiers in terms of performance measures. The ensemble classifier combined the classifiers such as NN, SVM, RF, and KNN in order to forecast these unseen examples, which depend on the vote to produce the addictive human predictions. The suggested LIME feature outperformed the current CNN, which obtained 59.33% accuracy, in the comparison, achieving 97.54% accuracy.

Table 4. Comparative analysis

Datasets	Method	Accuracy (%)	Precision (%)	F-score (%)	Standard deviation
Ecological momentary assessment [25]	Component-wise gradient boosting (CGB)	80	-	-	-
Collected data from instagram [26]	LSTM	-	68.6	72.4	-
Twitter [27]	CNN	59.33	95.4	67.7	-
Collected drug consumption data [28]	Ensemble model	93.33	93	93	-
Zengcheng compulsory isolated detoxification cente [29]	Gradient boosting regression tree (GBRT)	-	-	-	8.38
Drug rehabilitation center in zhejiang province [30]	PCA-k-means clustering	-	-	-	1.2
Collected drug consumption data	Proposed LIME with ensemble model approach	97.54	97.05	96.05	-

5. CONCLUSION

The LIME is used for finding the input features that are important for the generation of particular output based on the decision. The ensemble classifier learns not only a single classifier features but with the group of classifiers that are ensembled to combine for performing the predictions of the classification of the unseen instances based on the voting. The proposed LIME feature with the ensemble classifier finds out the input features of human behavior that are most important generates a particular output based on decision service. Since all of these unseen instances are relied on voting to make the human addiction predictions, the ensemble classifier included classifiers such NN, SVM, RF, and KNN to perform the prediction. Compared to the current CNN, which obtained 59.33% accuracy, and the ensemble model, which obtained 93.33% accuracy, the suggested LIME feature with the ensemble classifier obtained 97.54% accuracy better. The suggested technique will be expanded to include social media sites like Facebook and Twitter in order to enhance behavioural addiction forecasts.

ACKNOWLEDGEMENTS

Thanks to Management and Department of CSE, Vel Tech High Tech Dr.Rangarajan Dr.Sakunthala Engineering College.

REFERENCES

[1] L. Hitcham, H. Jackson, and R. J.E. James, “The relationship between smartphone use and smartphone addiction: an examination of logged and self-reported behavior in a pre-registered, two-wave sample,” *Computers in Human Behavior*, vol. 146, p. 107822, Sep. 2023, doi: 10.1016/j.chb.2023.107822.

[2] U. Granziol and G. Cona, “Architecture and relationships among cognition, mental health and other human domains revealed by network analysis perspective,” *Current Psychology*, May 2023, doi: 10.1007/s12144-023-04698-8.

[3] M. N. Rivera-Ríos, J. J. Cabiya, and I. Sánchez-Cardona, “Risk and protective factors predicting prescription drug misuse in a sample of Puerto Rican students,” *Journal of Addictive Diseases*, vol. 40, no. 1, pp. 78–83, Jan. 2022, doi: 10.1080/10550887.2021.1930825.

[4] J.-H. Ye *et al.*, “Predicting the learning avoidance motivation, learning commitment, and silent classroom behavior of chinese vocational college students caused by short video addiction,” *Healthcare*, vol. 11, no. 7, p. 985, Mar. 2023, doi: 10.3390/healthcare11070985.

[5] J. K. Brynildsen *et al.*, “Gene coexpression patterns predict opiate-induced brain-state transitions,” *Proceedings of the National Academy of Sciences*, vol. 117, no. 32, pp. 19556–19565, Aug. 2020, doi: 10.1073/pnas.2003601117.

[6] B. Harerimana, M. Kerr, R. Csiernik, L. C. Ng, E. Rutembesa, and C. Forchuk, “Predicting the contribution of age at first substance use and post-traumatic stress disorder to later addiction severity in a clinical sample from sub-saharan Africa: implications for prevention and treatment,” *International Journal of Mental Health and Addiction*, vol. 20, no. 3, pp. 1370–1384, Jun. 2022, doi: 10.1007/s11469-020-00449-z.

[7] K. Amasyali and N. El-Gohary, “Machine learning for occupant-behavior-sensitive cooling energy consumption prediction in office buildings,” *Renewable and Sustainable Energy Reviews*, vol. 142, p. 110714, May 2021, doi: 10.1016/j.rser.2021.110714.

[8] V. H. Pham, Q. H. Nguyen, T. T. Le, T. X. D. Nguyen, and T. T. K. Phan, “A proposal model using deep learning model integrated with knowledge graph for monitoring human behavior in forest protection,” *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 20, no. 6, p. 1276, Dec. 2022, doi: 10.12928/telkomnika.v20i6.24087.

[9] Y. Zhang *et al.*, “Towards augmented kernel extreme learning models for bankruptcy prediction: algorithmic behavior and comprehensive analysis,” *Neurocomputing*, vol. 430, pp. 185–212, Mar. 2021, doi: 10.1016/j.neucom.2020.10.038.

[10] P. Tripicchio and S. D’Avella, “Modeling multiple vehicle interaction constraints for behavior prediction of vehicles on highways,” *Computers & Electrical Engineering*, vol. 98, p. 107700, Mar. 2022, doi: 10.1016/j.compeleceng.2022.107700.

[11] S. Ghosh, P. Shivakumara, P. Roy, U. Pal, and T. Lu, “Graphology based handwritten character analysis for human behaviour identification,” *CAAI Transactions on Intelligence Technology*, vol. 5, no. 1, pp. 55–65, Mar. 2020, doi: 10.1049/trit.2019.0051.

[12] N. Parent, T. Bond, A. Wu, and J. Shapka, “Predicting patterns of problematic smartphone use among University students: a latent class analysis,” *Human Behavior and Emerging Technologies*, vol. 2022, pp. 1–12, Aug. 2022, doi: 10.1155/2022/4287600.

[13] K. A. Alissa *et al.*, "Appling tracking game system to measure user behavior toward cybersecurity policies," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 12, no. 5, p. 5164, Oct. 2022, doi: 10.11591/ijece.v12i5.pp5164-5175.
[14] İ. Hatun and N. B. Serin, "Positive conflict resolution behaviors in psychological health: Northern Cyprus sample," *International Journal of Public Health Science (IJPHS)*, vol. 10, no. 4, p. 865, Dec. 2021, doi: 10.11591/ijphs.v10i4.21007.
[15] F. Gao, B. Macquaire, Y. Zhang, and M. Bellet, "A new localized inverse identification method for high temperature testing under resistive heating: application to the elastic-viscoplastic behaviour of L-PBF processed In718," *Strain*, vol. 58, no. 3, Jun. 2022, doi: 10.1111/str.12409.
[16] D. M. Liday and M. R. C. Liwag, "Eating behavior and physical activity of senior citizens during the COVID-19 lockdown," *International Journal of Public Health Science (IJPHS)*, vol. 10, no. 3, p. 493, Sep. 2021, doi: 10.11591/ijphs.v10i3.20827.
[17] H. Rong, A. P. Teixeira, and C. Guedes Soares, "Ship collision avoidance behaviour recognition and analysis based on AIS data," *Ocean Engineering*, vol. 245, p. 110479, Feb. 2022, doi: 10.1016/j.oceaneng.2021.110479.
[18] J. Idrais, Y. El Moudene, and A. Sabour, "Characterizing user behavior in online social networks: analysis of the regular use of Facebook," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 11, no. 4, p. 3329, Aug. 2021, doi: 10.11591/ijece.v11i4.pp3329-3337.
[19] P. R. Walsh and R. Dodds, "The impact of intermediaries and social marketing on promoting sustainable behaviour in leisure travellers," *Journal of Cleaner Production*, vol. 338, p. 130537, Mar. 2022, doi: 10.1016/j.jclepro.2022.130537.
[20] A. Kumar, "Economically poor buyer: a systematic review of consumer behaviour research confronting to FMCGs," *International Journal of Business Innovation and Research*, vol. 27, no. 4, p. 480, 2022, doi: 10.1504/IJBIR.2022.122492.
[21] H. Zhu and H. Ou, "Constitutive modelling of hot deformation behaviour of metallic materials," *Materials Science and Engineering: A*, vol. 832, p. 142473, Jan. 2022, doi: 10.1016/j.msea.2021.142473.
[22] G. Waśniewski, K. Schabowicz, K. Wróblewski, and T. Kasprzak, "Identification of physical model of resinous material filling expansion joint in reinforced concrete structures," *Journal of Building Engineering*, vol. 45, p. 103505, Jan. 2022, doi: 10.1016/j.jobe.2021.103505.
[23] J. Xu *et al.*, "Global urbanicity is associated with brain and behaviour in young people," *Nature Human Behaviour*, vol. 6, no. 2, pp. 279–293, Oct. 2021, doi: 10.1038/s41562-021-01204-7.
[24] R. Baring, F. Magno, S. Reysen, and I. Katzarska-Miller, "Student religious attitudes and global citizenship antecedents, identification, and outcomes in a Filipino sample," *Journal of Beliefs & Values*, vol. 43, no. 2, pp. 159–170, Apr. 2022, doi: 10.1080/13617672.2021.1926758.
[25] S. T. Walters, M. S. Businelle, R. Suchting, X. Li, E. T. Hébert, and E.-Y. Mun, "Using machine learning to identify predictors of imminent drinking and create tailored messages for at-risk drinkers experiencing homelessness," *Journal of Substance Abuse Treatment*, vol. 127, p. 108417, Aug. 2021, doi: 10.1016/j.jsat.2021.108417.
[26] S. Hassanpour, N. Tomita, T. DeLise, B. Crosier, and L. A. Marsch, "Identifying substance use risk based on deep neural networks and Instagram social media data," *Neuropsychopharmacology*, vol. 44, no. 3, pp. 487–494, Feb. 2019, doi: 10.1038/s41386-018-0247-x.
[27] J. Tassone, P. Yan, M. Simpson, C. Mendhe, V. Mago, and S. Choudhury, "Utilizing deep learning and graph mining to identify drug use on Twitter data," *BMC Medical Informatics and Decision Making*, vol. 20, no. S11, p. 304, Dec. 2020, doi: 10.1186/s12911-020-01335-3.
[28] A. Shahriar, F. Faisal, S. U. Mahmud, A. Chakrabarti, and M. G. Rabiul Alam, "A machine learning approach to predict vulnerability to drug addiction," in *2019 22nd International Conference on Computer and Information Technology (ICCIT)*, IEEE, Dec. 2019, pp. 1–7. doi: 10.1109/ICCIT48885.2019.9038605.
[29] H. Lu, C. Xie, P. Lian, C. Yu, and Y. Xie, "Psychosocial factors predict the level of aggression of people with drug addiction: a machine learning approach," *Psychology, Health & Medicine*, vol. 27, no. 5, pp. 1168–1175, May 2022, doi: 10.1080/13548506.2021.1910321.
[30] Y. Yuan, J. Huang, and K. Yan, "Virtual reality therapy and machine learning techniques in drug addiction treatment," in *2019 10th International Conference on Information Technology in Medicine and Education (ITME)*, IEEE, Aug. 2019, pp. 241–245. doi: 10.1109/ITME.2019.00062.
[31] M. R. Haider, M. J. Brown, R. D. Gupta, S. Karim, B. Olatosi, and X. Li, "Psycho-social correlates of opioid use disorder among the US adult population: evidence from the national survey on drug use and health, 2015–2018," *Substance Use & Misuse*, vol. 55, no. 12, pp. 2002–2010, Sep. 2020, doi: 10.1080/10826084.2020.1788086.
[32] Y. Zhang, Z. Jiang, C. Chen, Q. Wei, H. Gu, and B. Yu, "DeepStack-DTIs: predicting drug–target interactions using lightGBM feature selection and deep-stacked ensemble classifier," *Interdisciplinary Sciences: Computational Life Sciences*, vol. 14, no. 2, pp. 311–330, Jun. 2022, doi: 10.1007/s12539-021-00488-7.
[33] S. G. Khalid, S. M. Ali, H. Liu, A. G. Qurashi, and U. Ali, "Photoplethysmography temporal marker-based machine learning classifier for anesthesia drug detection," *Medical & Biological Engineering & Computing*, vol. 60, no. 11, pp. 3057–3068, Nov. 2022, doi: 10.1007/s11517-022-02658-1.
[34] E. H. Houssein, M. E. Hosney, D. Oliva, W. M. Mohamed, and M. Hassaballah, "A novel hybrid Harris hawks optimization and support vector machines for drug design and discovery," *Computers & Chemical Engineering*, vol. 133, p. 106656, Feb. 2020, doi: 10.1016/j.compchemeng.2019.106656.
[35] A. K. S. Ong *et al.*, "Utilization of random forest classifier and artificial neural network for predicting factors influencing the perceived usability of COVID-19 contact tracing 'MorChana' in Thailand," *International Journal of Environmental Research and Public Health*, vol. 19, no. 13, p. 7979, Jun. 2022, doi: 10.3390/ijerph19137979.
[36] Y. Zhang, Y. Qiu, Y. Cui, S. Liu, and W. Zhang, "Predicting drug-drug interactions using multi-modal deep auto-encoders based network embedding and positive-unlabeled learning," *Methods*, vol. 179, pp. 37–46, Jul. 2020, doi: 10.1016/j.ymeth.2020.05.007.
[37] C. Sun, P. Xuan, T. Zhang, and Y. Ye, "Graph convolutional autoencoder and generative adversarial network-based method for predicting drug-target interactions," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 19, no. 1, pp. 455–464, Jan. 2022, doi: 10.1109/TCBB.2020.2999084.
[38] H. Zhang, J. Mao, H.-Z. Qi, and L. Ding, "In silico prediction of drug-induced developmental toxicity by using machine learning approaches," *Molecular Diversity*, vol. 24, no. 4, pp. 1281–1290, Nov. 2020, doi: 10.1007/s11030-019-09991-y.
[39] F. Rayhan, S. Ahmed, Z. Mousavian, D. M. Farid, and S. Shatabda, "FRnet-DTI: deep convolutional neural network for drug-target interaction prediction," *Heliyon*, vol. 6, no. 3, p. e03444, Mar. 2020, doi: 10.1016/j.heliyon.2020.e03444.
[40] C. Chen *et al.*, "DNN-DTIs: improved drug-target interactions prediction using XGBoost feature selection and deep neural network," *Computers in Biology and Medicine*, vol. 136, p. 104676, Sep. 2021, doi: 10.1016/j.compbiomed.2021.104676.

BIOGRAPHIES OF AUTHORS

V Sabapathi currently working as an Assistant Professor at Vel Tech High Tech Dr. Rangarajan Dr. Sakunthala Engineering College. His research area is cognitive systems, machine learning, deep learning, human behavioral analysis, data mining, and data analytics. He has published and presented papers in various international and national journals/conferences. He can be contacted at email: sabapathi@velhightech.com.

Selvin Paul Peter Jacob is an Associate Professor in the Department of Computer Science and Engineering at SRM Institute of Science and Technology, Kattankulathur. His area of interest is cloud computing and the internet of things (IoT). He has vast experience in academia and Industry. He has published and presented papers in various international and national journals/conferences. He is an active member of professional bodies such as the Institution of Engineers and the Indian Science Congress Association (ISCA). He can be contacted at email: selvinpj@srmist.edu.in.

Woothukadu Thirumaran Chembian received the bachelor's degree in Electronics from the University of Madras, Tamil Nadu, India, in 1997, M.E Degree from Anna University, Chennai, Tamil Nadu, India in 2007 and Ph.D. in Computer Science and Engineering from Annamalai University, Chidambaram, Tamil Nadu, India in 2020. His research interest includes, image mining, neural networks, back propagation deep learning and artificial intelligence. From 2000 to 2020, he was worked various Anna University Affiliated Institutions. He is currently a full time Associate Professor in the Department of Computer Science and Engineering, Vel Tech High Tech Dr. Rangarajan Dr. Sakunthala Engineering College (Autonomous), 60, Avadi – Vel Tech Road Vel Nagar Avadi, Chennai, Tamil Nadu. He can be contacted at email: wtchembian@velhightech.com.

Kandasamy Thinakaran received Ph.D. degree in computer science from Anna University, Tamilnadu in 2017. He is currently working an Associate Professor in Computer Science Engineering, Saveetha School of Engineering, Chennai India. His current research interests include neural network and data mining. He can be contacted at email: thinakarank.sse@saveetha.com.