Indonesian Journal of Electrical Engineering and Computer Science
Vol. 38, No. 2, May 2025, pp. 1231∼1244
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v38.i2.pp1231-1244

Textual and numerical data fusion for depression detection: a machine learning framework

Mohammad Tarek Aziz[1], Tanjim Mahmud[2], Md Faisal Bin Abdul Aziz[3], Md Abu Bakar Siddick[4], Md. Maskat Sharif[1], Mohammad Shahadat Hossain[5,6], Karl Andersson[6]

[1]Department of Computer Science and Engineering, Chittagong University of Engineering and Technology, Chittagong, Bangladesh
[2]Department of Computer Science and Engineering, Rangamati Science and Technology University, Rangamati, Bangladesh
[3]Department of Computer Science and Engineering, Comilla University, Comilla, Bangladesh
[4]Department of Computer Science and Technology, Beijing Institute of Technology, Haidian, China
[5]Department of Computer Science and Engineering, University of Chittagong, Chittagong, Bangladesh
[6]Cybersecurity Laboratory, Luleå University of Technology, Skellefteå, Sweden

Article Info

Article history:

Received Mar 28, 2024
Revised Oct 13, 2024
Accepted Oct 30, 2024

Keywords:

Chit-squared test
COVID-19
Depression
Machine learning algorithms
Random forest

ABSTRACT

Depression, a widespread mood disorder, significantly affects global mental health. To mitigate the risk of recurrence, early detection is crucial. This study explores socioeconomic factors contributing to depression and proposes a novel machine learning (ML)-based framework for its detection. We develop a tailored questionnaire to collect textual and numerical data, followed by rigorous feature selection using methods like backward removal and Pearson's chi-squared test. A variety of ML algorithms, including random forest (RF), support vector machine (SVM), and logistic regression (LR), are employed to create a predictive classifier. The RF model achieves the highest accuracy of 96.85%, highlighting its effectiveness in identifying depression risk factors. This research advances depression detection by integrating socioeconomic analysis with ML, offering a robust tool for enhancing predictive accuracy and enabling proactive mental health interventions.

Corresponding Authors:

Tanjim Mahmud
Rangamati Science and Technology University, Rangamati 4500, Bangladesh
Email: tanjim_cse@yahoo.com

1. INTRODUCTION

The rapid advancement of technology and human skills has profoundly influenced both physical and mental health [1]. Despite the growing awareness of physical well-being, mental health often remains undervalued, leading to a rise in mental health issues that evolve into disorders, including depression. Depression, a prevalent and debilitating mental illness, affects individuals of all ages, impairing their emotions, thoughts, and behaviors [2]. In the 2019–2020 academic year alone, a staggering 20.78% of adults experienced mental health issues, with over 50 million Americans affected [3]. The global COVID-19 pandemic has further exacerbated mental health challenges, with anxiety and hopelessness increasing by 25% worldwide [4]. Even post-recovery from the virus, individuals may still face lingering effects contributing to depression [5].

The pandemic's multifaceted impact, spanning mental, physical, and economic realms, has manifested in heightened levels of anxiety, insomnia, substance abuse, and other adverse outcomes. Socioeconomic factors such as unemployment, familial disconnection, and social isolation have compounded feelings of despair and disillusionment [6]. Alarmingly, approximately 280 million individuals worldwide suffer from depression, with a significant portion receiving inadequate treatment, particularly in low- and middle-income countries.

The global suicide rate, a tragic consequence of untreated depression, underscores the urgency of addressing mental health issues. In the realm of depression research, several studies have tackled various aspects of mental health detection using machine learning (ML) techniques.

Ferdowsy *et al.* [7] focused on predicting obesity risk, employing nine ML algorithms on a dataset comprising 1,100 instances and 28 features. Logistic regression (LR) emerged with the highest accuracy of 97.09%, while gradient boosting (GB) classifier yielded the lowest accuracy at 64.08%. Feature selection techniques such as correlation and principal component analysis (PCA) were utilized to identify critical features. Arif *et al.* [8] proposed a ML technique for detecting drug addiction in the population of Bangladesh. Nine ML classifiers were applied to a dataset containing 510 instances and 23 attributes, with LR achieving the highest accuracy of 97.91%. Correlation and PCA were also employed for feature selection to enhance model performance. Khatun *et al.* [9] investigated the detection of betel nut addiction using ML techniques. Among six classifiers applied to a dataset comprising 1001 samples and 19 features, random forest (RF) achieved the highest accuracy of 99.00%, while Naive Bayes (NB) obtained the lowest accuracy at 91.04%. Feature selection methods, including PCA and chi-square, were employed for improved accuracy. Mia *et al.* [10] introduced a ML approach to determine the registration status of students in private universities in Bangladesh. Seven ML algorithms were applied to a dataset with five features, achieving an accuracy of 85.76% with support vector machine (SVM) and 79.65% with RF.

Shahriar *et al.* [11] aimed to predict vulnerability to drug addiction using ML algorithms. Three classifiers were applied to a dataset comprising 498 samples and 60 attributes, with RF achieving the highest accuracy of 94.00%. Lee and Kim [12] focused on predicting problematic smartphone use using ML techniques. Among three classifiers applied to a dataset with 29,712 instances and 27 features, RF achieved the highest accuracy of 82.59%, while decision tree (DT) had the lowest accuracy at 74.56%. Keya *et al.* [13] employed ML approaches to analyze the performance of garment women's working status in Bangladesh. Five ML algorithms were applied to a dataset with 512 instances and 13 features, with LR achieving the highest accuracy of 69%. Govindasamy and Palanichamy [14] utilized ML techniques on Twitter data to detect depression. Two classifiers were applied to datasets with 1,000 and 3,000 instances respectively, with both models achieving an accuracy of around 92.34% and 97.31%. Sadeque *et al.* [15] explored depression detection in social media by focusing on individuals exhibiting signs of depression within online communities. Amanat *et al.* [16] employed deep learning algorithms to predict depression, achieving an accuracy of 99% using long short-term memory (LSTM) and recurrent neural network (RNN) algorithms on textual data.

Against this backdrop, our research endeavors to contribute to the early detection of depression using ML algorithms [17], [18] and feature selection techniques [19]. By analyzing a dataset comprising 6,186 instances and 28 features related to depression. We aim to identify individuals at risk of depression and distinguish them from non-depressed individuals. The primary objectives of our study include detecting depression using mixed data (both numerical and textual), gathering comprehensive data through tailored questionnaires, and applying feature selection techniques to enhance model accuracy.

The main contributions of this study are below:

i) Development of a combined dataset incorporating both textual and numerical data, augmented through oversampling and chi-square technique to increase dataset robustness.
ii) Utilization of the developed dataset to train ML classifiers for accurate prediction of depression, achieved by converting dataset into binary values for labeling.
iii) Proposal of a RF model, demonstrating superior performance with training accuracy of 97.37% and testing accuracy of 96.85%, surpassing other classifiers.
iv) Application of nine ML classifiers to effectively detect depression from mixed data, achieving accuracy rates exceeding 78% in most cases.

In this study, the proposed system architecture in section 2, the result analysis in section 3, and finally the conclusion and future plan in section 4 are discussed sequentially.

2. METHOD

The working procedure of the proposed system architecture is illustrated in Figure 1, and an explanation of the mechanism is described in below.

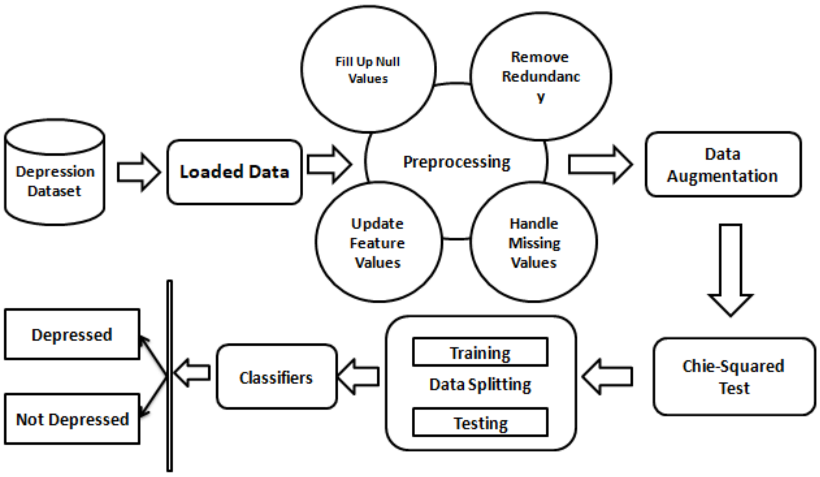

Figure 1. Proposed system architecture of depression detection

2.1. Dataset

Data collection was conducted through online sources including websites, web portals, and social media platforms such as Facebook, Instagram, and Twitter, utilizing the data-collecting tool Facepager. Additionally, face-to-face interviews were conducted with individuals affiliated with medical institutions, hospitals, and clinics. This approach involved interacting with individuals who were diagnosed with depression or exhibited symptoms of mental disorder, albeit being comparatively challenging and time-consuming. Despite the challenges, a total of 520 records were obtained through these two methods [20]. To comprehend the regional factors contributing to depression, a questionnaire was developed by reviewing literature, relevant websites, and consulting regional psychiatrists. Initially, a draft questionnaire was created, which was later refined to include a total of 28 items. These questions are structured as multiple choice questions (MCQs) with two or more possible solutions. A dataset comprising 520 cases was collected based on the responses to these questions, which serve as input data for the proposed system. Each question in the questionnaire presents respondents with specific answer options, such as binary choices like “yes” or “no”. The data collection process is depicted in Figure 2, encompassing various steps outlined below.

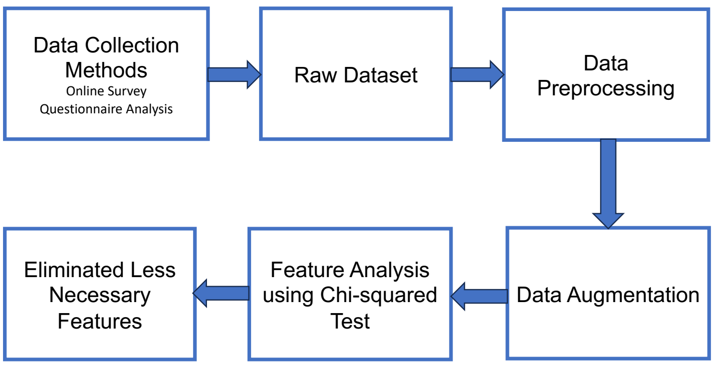

Figure 2. Working process of data collection and feature analysis with chi-squared test

2.1.1. Questionnaire analysis

We administered a set of 28 questions to users to gather data, with the intention of extracting information for future analysis based on their responses [21]. These questions, detailed in Table 1, encompass various aspects related to the individual’s circumstances, including inquiries about age, family crises, and mental pressure [22]. The solutions to these questions typically involve more than two options, such as “yes” or “no”, with each response contributing to the assessment of the individual’s likelihood of experiencing depression. For instance, questions pertaining to age, family dynamics, and mental well-being are particularly relevant indicators. Following the collection of responses, the model can then determine whether the individual exhibits signs of depression or not. For a comprehensive list of questions, please refer to Table 1, while Table 2 presents an analysis of the corresponding solutions.

Table 1. Questions to collect data from person interview or website making a dataset

S.L	User's information to find out depression
1	Your Age range?
2	Your gender?
3	Your occupation?
4	What kind of relationship is between you and your family?
5	Can you share your personal matter with your family members easily?
6	Do you have a family crisis?
7	Is your family supportive?
8	Do you have faced any Violence in your family?
9	How did you spend your lockdown time during COVID 19?
10	During COVID 19 lockdown, did you feel lonely?
11	If you are a student, are you satisfied with your academic result?
12	If you are a student, Online class during COVID 19 affect your mental health?
13	Did you retreat from your study because of COVID 19?
14	If you are an undergraduate, have you fallen in Session clutter because of COVID 19?
15	Are you afraid to apply abroad for your higher study because of COVID 19?
16	What kind of relationship is between you and your friends/ colleagues?
17	What is your relationship status?
18	Are you happy with your partner?
19	Are you tensed about anything that you cannot forget anymore?
20	Are you worried about the uncertainty of getting a job because of COVID 19?
21	Are you happy with your current situation?
22	Did you attack yourself for some reason during COVID 19?
23	Did you loss your closer anyone due to COVID 19 and you cannot forget this situation till now?
24	Which things are always pressurized on you?
25	Are you addicted to any drugs?
26	Have you ever experienced bullying from friends or through social media?
27	Do you spend most of your time on Social media?
28	Do you have sleeping (insomnia) problem?

Table 2. Top 10 features of collected raw dataset with top five value

Features	Data 1	Data 2	Data 3	Data 4	Data 5
Age	18-25	18-25	25-40	18-25	18-25
Gender	Female	Female	Male	Male	Male
Occupation	Student	Student	Service holder	Student	Student
Relationships	Strong	Strong	Normal	Strong	Strong
Data sharing	No	Yes	Yes	Yes	Yes
Family-crisis	No	No	No	No	No
Mental pressure	Yes	No	Yes	No	No
Supportive-family	Yes	Yes	Yes	Yes	Yes
Family-violence	No	No	No	No	Yes
Violence pressure	No	Yes	No	No	Yes

2.1.2. Raw dataset

After finishing the data collection, we got a total of 520 records with 28 features. The features are: 'Age', 'gender', 'occupation', 'relation-with-family', 'shared-personal-matter-with-family', 'family-crisis', 'mental-pressure-due-to-crisis', 'supportive-family', 'faced-Violence', 'mental-pressure-due-to-violence', 'spend-lockdown', 'feel-lonely-during-COVID19', 'Satisfied-academic-result', 'mental-pressure-due-to-online-class', 'retreat-from-study-due-to-COVID19', 'mental-pressure-due-to-retreat', 'fallen-into-session-clutter', 'mental-pressure-due-to-session-clutter', 'afraid-apply-abroad', 'relation-with-friends', 'relationship-status', 'happy-with-partner', 'tensed-something', 'worried-uncertainty-getting-job-COVID-19', 'mental-pressure-due-to-getting-job', 'happy-with-current-situation', 'attack-yourself-in-COVID-19-situation', 'lostcloser-during-COVID-19', 'pressurized-on-you', 'addicted', 'bullied-social media', 'mental-pressure-due-to-bullied', 'spend-much-time-social-media', 'insomnia-problem'. Only "age" contains the numerical value, and the other features contain the textual value. In most cases, textual-based features have two values: yes or no. The top 10 features are shown in Table 2 with their values. The other 18 features also have almost the same value, like data sharing and violent pressure. That means they contain two Boolean values: yes and no. So, we ignored adding these features here due to length.

2.2. Data labeling

Data labeling, also known as data annotation, plays a pivotal role in ML tasks [23]. Labeled data is essential to prevent issues such as overfitting or underfitting, which may occur when using unlabeled data. Without proper labeling, classifiers may misclassify or exhibit reduced accuracy, thereby disrupting the desired outcome [24].

In our proposed method, we opted for binary transformation as most features possess two distinct values, such as "yes" or "no". Table 3 displays the binary-transformed values for a total of 28 features, showcasing the top 5 values. While presenting the top 10 features in this table, we omitted the remaining features as they exhibit similar characteristics.

Table 3. Binary transformed value for some features with top five values

Features	Data 1	Data 2	Data 3	Data 4	Data 5
Relationships	1	1	0	1	1
Data sharing	0	1	1	1	1
Family-crisis	0	0	0	0	0
Mental pressure	1	0	1	0	0
Supportive-family	1	1	1	1	1
Family-violence	0	0	0	0	1
Violence pressure	0	0	0	0	1

2.3. Data preprocessing

The questionnaire serves the purpose of gathering essential insights into the primary causes of depression prevalent in Bangladesh. Data collection was conducted through two methods: online search and personal interviews. Ultimately, a total of 520 instances were amassed. The subsequent step in our methodology involves data preprocessing, which encompasses four fundamental tasks [25], [26]:

1. Null values were filled electronically to ensure completeness of the dataset.
2. Missing values labeled as "nan" were manually addressed to ensure compatibility with subsequent operations.
3. Some feature values were updated to binary "yes" or "no" format to align with the classification method to be used in the model.
4. Redundant values were eliminated from the dataset to streamline the analysis process.

2.4. Data augmentation

Given the imbalanced nature of our dataset, we employed oversampling techniques [27], [28] to rectify the imbalance. A commonly used method in the literature for generating additional samples is the synthetic minority over-sampling technique (SMOTE). Through oversampling, our dataset expanded by approximately 6,186 samples, resulting in a well-balanced dataset suitable for training our ML model. Specifically, Table 4 showcases a selection of seven randomly chosen features along with their corresponding matrix format, following oversampling. Other features were disregarded due to their similarity.

Table 4. Over sampled matrix format of the features

Feature	Matrix format
Gender	(610, 26)
shared-personal Matter_-with-family	(702, 26)
Supportive-family	(1100, 26)
Mental Pressure-due-to-violence	(1592, 26)
Addicted	(5846, 26)
Insomnia-problem	(6258, 26)
Attack-yourself-in-COVID19_situation	(2948, 26)

2.5. Chi-squared test analysis and feature selection

Without assuming any particular distribution, the Pearson's chi-squared test of independence is a statistical technique used to evaluate the relationship between two categorical variables based on their frequencies [29]. This test enables us to determine whether there is a significant correlation between the predictor variables and the target variable in our dataset, where all of the variables are categorical. This test's null hypothesis presupposes that there is no relationship between the variables.

The chi-squared test, shown in Figure 3, can be used to ascertain whether the target variable and the feature variables are dependent on one another. This is accomplished by creating a contingency table, which arranges the variable frequencies in an organized manner and is also referred to as a cross-tabulation or two-way table. Based on the variations between observed and expected frequencies in the contingency table, the test computes a chi-squared statistic [30]. We reject the null hypothesis and conclude that there is a significant association between the variables if the computed chi-squared statistic is greater than a critical value determined by the degrees of freedom and selected significance level. In contrast, if the computed statistic is less than the critical value, we are unable to reject the null hypothesis, indicating that there is no meaningful correlation.

The analytical form of the chi-squared test statistic for a contingency table with r rows and c columns is given by:

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

where:

- O_{ij} represents the observed frequency in the ith row and jth column of the contingency table.
- E_{ij} represents the expected frequency in the ith row and jth column, calculated under the assumption of independence between the two categorical variables.

The chi-squared test statistic follows a chi-squared distribution with degrees of freedom equal to $(r-1)(c-1)$ under the null hypothesis of independence between the variables. By comparing the calculated chi-squared statistic with the critical value from the chi-squared distribution, one can determine whether to reject or fail to reject the null hypothesis. Using the chi-square test, based on the relationship between each independent feature and target variable, we selected 25 variables. For one independent variable, the null hypothesis was false. So we removed this feature. Among the original 26 features in our data collection, the 25 special variables used for creation have been evaluated as having the highest impact.

```
[[  74   51]
 [1987 1230]
 [2194  611]
 [  22   89]]
Age vs Family-crisis

[[1128  656]
 [1371  962]
 [1700  323]
 [  78   40]]
Occupation vs Family-
crisis

[[2333  843]
 [1944 1138]]
Faced-violence vs
Family-crisis

[[ 794  451]
 [ 366  444]
 [ 183  117]
 [2934  969]]
Relationship-status
vs Family-crisis
```

Figure 3. Feature analysis with chi-test for target and other variables

2.6. Classifiers

In this study, we applied a total of 9 ML classifiers to detect depression from both textual and numerical datasets. The classifiers are used to find out the depression with optimum solutions. The used classifiers are described below.

2.6.1. Random forest

RF is a popular ensemble learning algorithm that constructs multiple DTs during training and combines their predictions to reduce overfitting and improve accuracy for both regression and classification tasks [31], [32]. Figure 4 shows working process of the proposed RF classifiers.

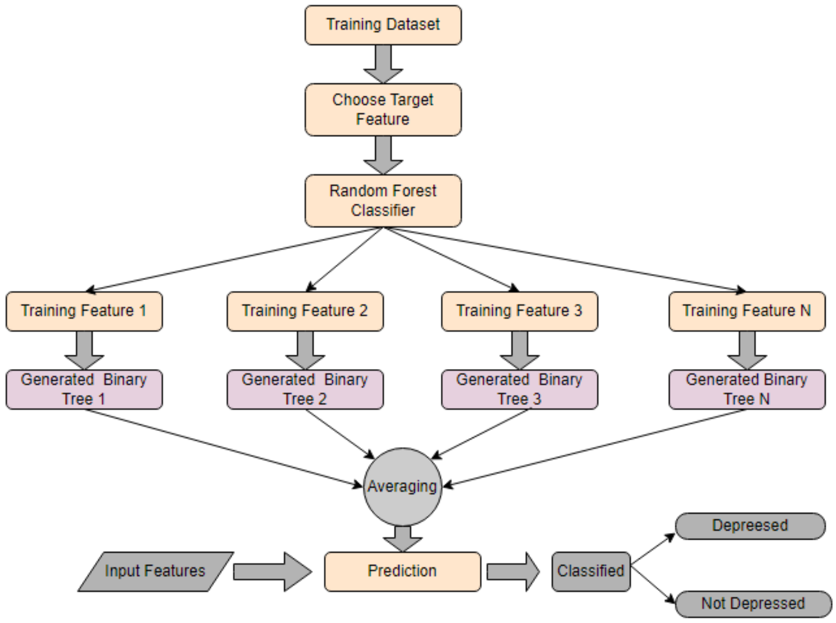

Figure 4. Working process of the proposed RF classifiers

2.6.2. Multilayer perceptron

The multilayer perceptron (MLP) is a basic artificial neural network (ANN) with a layered architecture consisting of an input layer, one or more hidden layers, and an output layer, enabling it to model complex non-linear relationships between inputs and outputs.

2.6.3. Decision tree

The DT algorithm is a popular and straightforward ML method that constructs a tree-like structure for regression and classification tasks, offering ease of use, interpretability, and the ability to handle both categorical and numerical data.

2.6.4. Logistic regression

LR is a statistical technique used for binary classification problems, modeling the probability of a binary outcome based on predictor variables and converting raw predictions into probabilities using the logistic function [33].

2.6.5. Gradient boosting

GB is an ensemble learning technique that builds a strong predictive model by iteratively adding weak learners, typically DTs, each improving on the residuals of the previous models to enhance overall accuracy.

2.6.6. Gaussian Naive Bayes

Gaussian NB is a probabilistic classification algorithm based on Bayes' theorem, assuming features are independent and normally distributed, making it effective for tasks like text classification and spam filtering, especially with continuous data.

2.6.7. Support vector machine

SVM is a powerful supervised learning algorithm that determines the optimal hyperplane to separate classes in the feature space, useful for both linear and non-linear data, particularly in high-dimensional spaces [34].

2.6.8. K-nearest neighbors

K-nearest neighbors (KNN) is a non-parametric ML algorithm that assigns a data point to the majority class among its nearest neighbors, commonly used for tasks with non-linear decision boundaries [35].

2.6.9. AdaBoost

AdaBoost (ADA-B), or adaptive boosting, is an ensemble technique that combines several weak learners by adjusting their weights based on misclassified instances, enhancing classification performance through iterative improvement.

3. RESULTS AND DISCUSSION

3.1. Data splitting

The dataset underwent a division into training and testing subsets, allocated at an 80% to 20% ratio, respectively. Both subsets utilized binary expressed values for training and testing purposes. Various ML techniques were employed to construct a predictive binary classifier using the training set. These techniques included GB, ADA-B classifier, KNN, gaussian NB, SVM, DT, RF, and LR. The selected target variable for the classifier was "Family crisis," chosen due to the significant role of financial issues in male depression. The model's output was compared against the target variable of the test set to evaluate the accuracy and effectiveness of the classifier.

3.2. Used classifier parameters

We review the many parameters utilized to build the classifiers and run experiments in this part. The parameters vary based on the classifiers. See the parameter details in Table 5 that are used in the study.

Table 5. Classifier parameter details

Classifiers	Parameter
LR	Dual=False, solver='lbfgs', penalty='l2'
DT	threshold='gini', maximum depth=25, random state=0 splitter = "optimal"
RF	The smallest samples are "4 for departs 9 for splits in half, 114 for estimations, and 22 for random states," while the maximum depth is 25 and the maximum features are "log2."
SVM	random state=0, kernel='rbf'
KNN	weights='uniform'
Gaussian NB	varsmoothing=1e-9
MLP	Random state = 0, activation = logistic, solver = lbfgs
GB classifiers	Loss = "deviance," learning rate = 0.1, and n estimators = 100
ADA-B classifier	n estimators= 50

3.3. Evaluation metric

In this study, we used classifiers to predict sad individuals. We used the Scikit-learn function accuracy score() [36], [37], which uses the test dataset's classifiers' projected results, to determine the accuracy.
TP (true positive): it indicates the amount of accurately predicted dataset by the classifier.
FP (false positive): the amount of dataset had the classifier incorrectly predicted as being in good health when they are depressed.
FN (false negative): the classifier mislabeled a number of outcomes as having addictions when, in fact, they were healthy in the sample.
TN (true negative): the number of outcomes in the dataset is correctly classified as depressed by the classifier.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} = \frac{2 * TP}{2 * TP + FP + FN} \tag{4}$$

In addition to figuring out these numbers, we also measured the AUC value and produced the ROC curve (receiver operator characteristic). In a ROC diagram, the X- and Y-axes for the TP rate (recall) and

FP rate (1-TNR) are displayed, respectively. Comparing the potency of different prediction models is done using the two-dimensional area under the ROC curve, or AUC. A higher AUC value typically indicates better prediction performance. The classifier that performs the best is chosen after taking into account all of these variables.

3.4. Confusion matrix analysis

If the explanation of error analysis were used, it would be simpler to comprehend how well the chosen classifier performed. A confusion matrix makes error analysis easier to understand. The value of it, the best classifier after training with 28 features is displayed in Figure 5. Out of the 1,238 total predictions, 848 respondents were correctly recognized as being in good health, while 17 healthy respondents were misclassified as being in bad health. In the instance of depression predictions, 41 depressed respondents were mistakenly classed as healthy, while the remaining 346 respondents were accurately recognized. Compared to depression responses, the classifier makes fewer mistakes in the healthy class.

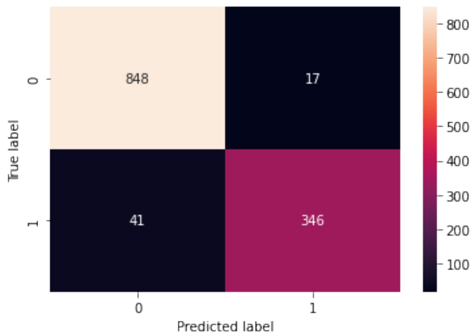

Figure 5. Confusion matrix of RF classifier

3.5. Accuracy scores

Since the feature analysis is performed using the chi-squared test, and we selected 28 features based on it, we performed nine algorithms shown in Table 6 and analyzed their accuracy score. We got the highest accuracy in DT classifiers at 99.49% for training and 96.44% for testing accuracy.

3.6. ROC and AUC value analysis

A classifier, named RF, has the best accuracy among trained classifiers, according to an analysis of the data in Table 6. We need to look at the ROC signal for every classifier to test performance in terms of sensitivity and specificity in addition to classification accuracy. Comparing classifiers was done using AUC values and ROC curves, which provide reliable descriptions of discriminating skills. Accuracy ratings may not fully explain the situation, thus AUC values are necessary to comprehend how the classifiers performed in both the healthy and addicted classes in our study. By comparing accuracy scores and AUC values, we may arrive at a convincing decision on which model is best. The results presented in Table 6 and Figure 6 show that the RF classifier has the highest correlation when we look at the AUC result for the learners who have the highest accuracy scores, with a success rate of 96.85% and a mean AUC of 0.99. Figure 7 shows ROC curves for SVM, KNN, and ADA-B classifiers.

Table 6. Training and testing accuracy of the used classifiers

Classifier	Training accuracy %	Testing accuracy %
LR	85.79	87.80
DT	99.49	96.44
RF	97.37	96.85
SVM	95.98	95.64
KNN	97.33	96.20
Gaussian NB	78.48	79.89
MLP	99.49	96.61
GB classifiers	92.92	93.38
ADA-B classifier	86.50	88.61

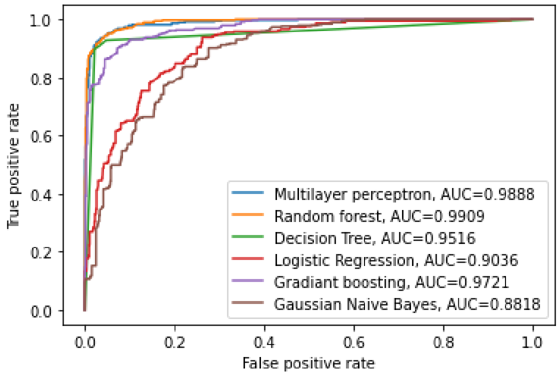

Figure 6. ROC curves for MLP, RF, DT, LT, GB, and gaussian NB classifiers

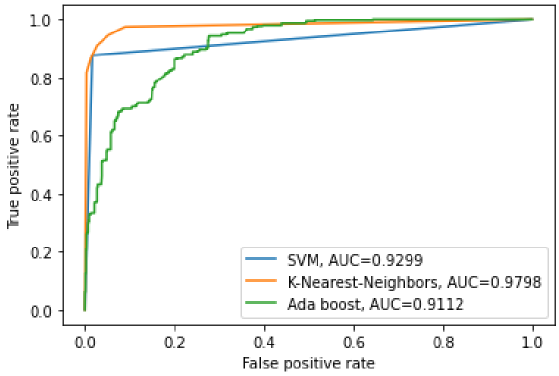

Figure 7. ROC curves for SVM, KNN, and ADA-B classifiers

3.7. Precision, recall, and F1-score analysis of the used classifiers

The classifier with the best accuracy and AUC value among all of them is a RF classifier trained with 28 feature variables. Consider Table 7, which contains explanations of each of the classifiers mentioned above along with additional metrics like recall and precision [38], [39].

Table 7. Precision, recall, and F1-score value for all classifier

Classifier	Class	Precision	Recall	F1-score
LR	Healthy	0.90	0.92	0.91
	Depressed	0.82	0.78	0.80
DT	Healthy	0.97	0.97	0.97
	Depressed	0.94	0.94	0.94
RF	Healthy	0.97	0.99	0.98
	Depressed	0.97	0.92	0.95
SVM	Healthy	0.96	0.98	0.97
	Depressed	0.96	0.90	0.93
KNN	Healthy	0.97	0.98	0.97
	Depressed	0.94	0.93	0.94
Gaussian NB	Healthy	0.91	0.79	0.84
	Depressed	0.63	0.82	0.72
MLP	Healthy	0.97	0.98	0.98
	Depressed	0.95	0.94	0.94
GB	Healthy	0.94	0.97	0.95
	Depressed	0.93	0.85	0.89
ADA-B	Healthy	0.90	0.94	0.92
	Depressed	0.85	0.76	0.80

3.8. Comparison with state-of-art method

The quality of our suggested depression prediction method can be assessed by comparing it to the work of some of our relatives. After reading a significant number of research articles, we discovered that there has been a lot of work done on prediction, but it is challenging to locate work that is pertinent to ours. Some work has been done mentioned in Table 8.

Table 8. Comparison with previous studies

Reference	Technique	Accuracy
[40]	ADA-B classifier	92.56%
[41]	SVM	78.59%
[42]	RF	75%
[43]	Convolutional neural network (CNN), RNN	87%, 80%
[5]	SVM, NB	79%, 83%
[4]	Gaussian mixture, SVM	70%, 75%
[44]	SVM, KNN, ANN	79.27%
[45]	Visual geometry group (VGG)-Net	90%
Our study	RF	96.85%

4. CONCLUSION

In this study, we present a robust framework for depression detection by integrating textual and numerical data using ML techniques. By analyzing socioeconomic factors and designing a tailored questionnaire, we identified key features contributing to depression and developed a predictive classifier using various ML algorithms. Our methodology, validated on real-world datasets, achieved a high accuracy of 96.85% with the RF model, highlighting its effectiveness in identifying depression risk factors. This research represents a significant advancement in depression detection strategies, providing a comprehensive framework for proactive intervention in mental health care. For future work, we aim to further enhance the predictive accuracy of the model by exploring additional feature selection techniques and optimizing the ML algorithms. Additionally, we plan to extend the research to incorporate more diverse datasets and validate the framework across different demographic groups and cultural contexts. Furthermore, integrating real-time monitoring and feedback mechanisms into the framework could enable timely intervention and personalized support for individuals at risk of depression.

ACKNOWLEDGEMENTS

The manuscript's grammar was also improved with an AI tool, which led to a notable improvement in its quality.

REFERENCES

[1] J. Sun, H. Shen, S. Ibn-ul-Hassan, A. Riaz, and A. E. Domil, "The association between digitalization and mental health: the mediating role of wellbeing at work," *Frontiers in Psychiatry*, vol. 13, p. 934357, 2022, doi: 10.3389/fpsyt.2022.934357.

[2] I. H. Gotlib and J. Joormann, "Cognition and depression: current status and future directions," *Annual review of clinical psychology*, vol. 6, pp. 285–312, 2010.

[3] M. R. Islam, M. A. Kabir, A. Ahmed, A. R. M. Kamal, H. Wang, and A. Ulhaq, "Depression detection from social network data using machine learning techniques," *Health Information Science and Systems*, vol. 6, no. 1, pp. 1–12, 2018, doi: 10.1007/s13755-018-0046-0.

[4] S. Alghowinem, R. Goecke, M. Wagner, G. Parker, and M. Breakspear, "Eye movement analysis for depression detection," in *2013 IEEE International Conference on Image Processing, ICIP 2013 - Proceedings*, 2013, pp. 4220–4224, doi: 10.1109/ICIP.2013.6738869.

[5] M. Deshpande and V. Rao, "Depression detection using emotion artificial intelligence," in *Proceedings of the International Conference on Intelligent Sustainable Systems, ICISS 2017*, 2018, pp. 858–862, doi: 10.1109/ISS1.2017.8389299.

[6] A. Jeffers *et al.*, "Impact of Social Isolation during the COVID-19 pandemic on mental health, substance use, and homelessness: qualitative interviews with behavioral health providers," *International Journal of Environmental Research and Public Health*, vol. 19, no. 19, p. 12120, 2022, doi: 10.3390/ijerph191912120.

[7] F. Ferdowsy, K. S. A. Rahi, M. I. Jabiullah, and M. T. Habib, "A machine learning approach for obesity risk prediction," *Current Research in Behavioral Sciences*, vol. 2, p. 100053, 2021, doi: 10.1016/j.crbeha.2021.100053.

[8] M. A. I. Arif, S. I. Sany, F. Sharmin, M. S. Rahman, and M. T. Habib, "Prediction of addiction to drugs and alcohol using machine learning: a case study on Bangladeshi population," *International Journal of Electrical and Computer Engineering*, vol. 11, no. 5, pp. 4471–4480, 2021, doi: 10.11591/ijece.v11i5.pp4471-4480.

[9] J. Khatun, T. Azad, S. Y. Seaum, M. I. Jabiullah, and M. T. Habib, "Betel nut addiction detection using machine learning," in *Proceedings of the 4th International Conference on Electronics, Communication and Aerospace Technology, ICECA 2020*, 2020, pp. 1435–1440, doi: 10.1109/ICECA49313.2020.9297437.

[10] M. J. Mia, A. A. Biswas, A. Sattar, and M. T. Habib, "Registration status prediction of students using machine learning in the context of private university of Bangladesh," *International Journal of Innovative Technology and Exploring Engineering*, vol. 9, no. 1, pp. 2594–2600, 2019, doi: 10.35940/ijitee.A5292.119119.

[11] A. Shahriar, F. Faisal, S. U. Mahmud, A. Chakrabarti, and M. G. R. Alam, "A machine learning approach to predict vulnerability to drug addiction," in *2019 22nd International Conference on Computer and Information Technology, ICCIT 2019*, 2019, pp. 1–7, doi: 10.1109/ICCIT48885.2019.9038605.

[12] J. Lee and W. Kim, "Prediction of problematic smartphone use: a machine learning approach," *International Journal of Environmental Research and Public Health*, vol. 18, no. 12, p. 6458, 2021, doi: 10.3390/ijerph18126458.

[13] M. S. Keya, M. U. Emon, H. Akter, M. A. M. Imran, M. K. Hassan, and M. U. Mojumdar, "Predicting performance analysis of garments women working status in Bangladesh using machine learning approaches," in *Proceedings of the 6th International Conference on Inventive Computation Technologies, ICICT 2021*, 2021, pp. 602–608, doi: 10.1109/ICICT50816.2021.9358696.

[14] K. A. L. Govindasamy and N. Palanichamy, "Depression detection using machine learning techniques on twitter data," in *Proceedings - 5th International Conference on Intelligent Computing and Control Systems, ICICCS 2021*, 2021, pp. 960–966, doi: 10.1109/ICICCS51141.2021.9432203.

[15] F. Sadeque, D. Xu, and S. Bethard, "Measuring the latency of depression detection in social media," in *WSDM 2018 - Proceedings of the 11th ACM International Conference on Web Search and Data Mining*, 2018, vol. 2018-Febuary, pp. 495–503, doi: 10.1145/3159652.3159725.

[16] A. Amanat *et al.*, "Deep learning for depression detection from textual data," *Electronics (Switzerland)*, vol. 11, no. 5, Mar. 2022, doi: 10.3390/electronics11050676.

[17] T. Mahmud, M. Ptaszynski, and F. Masui, "Deep learning hybrid models for multilingual cyberbullying detection: insights from Bangla and Chittagonian languages," in *2023 26th International Conference on Computer and Information Technology, ICCIT 2023*, 2023, pp. 1–6, doi: 10.1109/ICCIT60459.2023.10441181.

[18] M. T. Aziz *et al.*, "Machine learning-driven job recommendations: harnessing genetic algorithms," in *International Congress on Information and Communication Technology*, 2024, pp. 471–480.

[19] S. U. Habiba *et al.*, "Early prediction of chronic kidney disease using machine learning algorithms with feature selection techniques," in *Communications in Computer and Information Science*, 2024, vol. 2065 CCIS, pp. 224–242, doi: 10.1007/978-3-031-68639-9_14.

[20] M. S. Hosen, "The ICT on the education system and it's future prospects in Bangladesh," *AGPE The Royal Gondwana Research Journal of History*, vol. 3, no. 2, pp. 48–60, 2022.

[21] L. R. Harris and G. T. L. Brown, "Mixing interview and questionnaire methods: practical problems in aligning data," *Practical Assessment, Research and Evaluation*, vol. 15, no. 1, p. 1, 2010.

[22] A. Burcu, "A comparison of two data collecting methods: interviews and questionnaires," *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, vol. 18, pp. 1–10, 2000.

[23] N. A. Chowdhury *et al.*, "A novel approach to detect stroke from 2D images using deep learning," in *International Conference on Big Data, IoT and Machine Learning*, 2023, vol. 867 LNNS, pp. 239–253, doi: 10.1007/978-981-99-8937-9_17.

[24] M. T. Aziz, R. Sudheesh, R. D. C. Pecho, N. U. A. Khan, A. Ull, H. Era, and M. A. Chowdhury, "Calories burnt prediction using machine learning approach," *Current Integrative Engineering*, vol. 1, no. 1, pp. 29–36, 2023, doi: 10.59762/cie570390541120231031130323.

[25] A. S. Khan *et al.*, "Integrating BERT embeddings with SVM for prostate cancer prediction," in *Proceedings - 6th International Conference on Electrical Engineering and Information and Communication Technology, ICEEICT 2024*, 2024, pp. 574–579, doi: 10.1109/ICEEICT62016.2024.10534547.

[26] A. B. Aziz *et al.*, "YOLO-V4 based detection of varied hand gestures in heterogeneous settings," in *International Conference on Applied Intelligence and Informatics*, 2023, pp. 325–338.

[27] A. D. Bappy, T. Mahmud, M. S. Kaiser, M. S. Hossain, and K. Andersson, "A BERT-based chatbot to support cancer treatment follow-up," in *International Conference on Applied Intelligence and Informatics*, 2024, pp. 47–64, doi: 10.1007/978-3-031-68639-9_4.

[28] M. H. Imam *et al.*, "A transfer learning-based framework: MobileNet-SVM for efficient tomato leaf disease classification," in *Proceedings - 6th International Conference on Electrical Engineering and Information and Communication Technology, ICEEICT 2024*, 2024, pp. 693–698, doi: 10.1109/ICEEICT62016.2024.10534539.

[29] Y. Zhai, W. Song, X. Liu, L. Liu, and X. Zhao, "A chi-square statistics based feature selection method in text classification," in *Proceedings of the IEEE International Conference on Software Engineering and Service Sciences, ICSESS*, 2018, vol. 2018-November, pp. 160–163, doi: 10.1109/ICSESS.2018.8663882.

[30] H. Byeon, "Chi-square automatic interaction detection modeling for predicting depression in multicultural female students," *International Journal of Advanced Computer Science and Applications*, vol. 8, no. 12, 2017, doi: 10.14569/ijacsa.2017.081222.

[31] M. A. Rahman, M. Begum, T. Mahmud, M. S. Hossain, and K. Andersson, "Analyzing sentiments in eLearning: a comparative study of Bangla and Romanized Bangla text using transformers," *IEEE Access*, vol. 12, pp. 89144–89162, 2024, doi: 10.1109/ACCESS.2024.3419024.

[32] T. Mahmud, M. Ptaszynski, and F. Masui, "Exhaustive study into machine learning and deep learning methods for multilingual cyberbullying detection in Bangla and Chittagonian texts," *Electronics (Switzerland)*, vol. 13, no. 9, p. 1677, 2024, doi: 10.3390/electronics13091677.

[33] T. Mahmud, T. Akter, S. Anwar, M. T. Aziz, M. S. Hossain, and K. Andersson, "Predictive modeling in forex trading: a time series analysis approach," in *Proceedings - 2024 2nd International Conference on Inventive Computing and Informatics, ICICI 2024*, 2024, pp. 390–397, doi: 10.1109/ICICI62254.2024.00070.

[34] T. Mahmud, U. K. Das, S. R. Naher, M. S. Hossain, K. Andersson, and others, “Investigating the effectiveness of deep learning and machine learning for bangla poems genre classification,” in *2023 4th International Conference on Intelligent Technologies (CONIT)*, 2024, pp. 1–6.

[35] F. Tasnim, S. U. Habiba, T. Mahmud, L. Nahar, M. S. Hossain, and K. Andersson, “Protein sequence classification through deep learning and encoding strategies,” *Procedia Computer Science*, vol. 238, pp. 876–881, 2024, doi: 10.1016/j.procs.2024.06.106.

[36] P. Dey, T. Mahmud, M. S. Chowdhury, M. S. Hossain, and K. Andersson, “Human age and gender prediction from facial images using deep learning methods,” *Procedia Computer Science*, vol. 238, pp. 314–321, 2024, doi: 10.1016/j.procs.2024.06.030.

[37] P. Dey, T. Mahmud, K. M. Foysol, N. Sharmen, M. S. Hossain, and K. Andersson, “Hybrid deep transfer learning framework for humerus fracture detection and classification from x-ray images,” in *2024 4th International Conference on Intelligent Technologies, CONIT 2024*, 2024, pp. 1–6, doi: 10.1109/CONIT61985.2024.10626930.

[38] F. Khan, R. Mustafa, F. Tasnim, T. Mahmud, M. S. Hossain, and K. Andersson, “Exploring BERT and ELMo for Bangla spam SMS dataset creation and detection,” in *2023 26th International Conference on Computer and Information Technology, ICCIT 2023*, 2023, pp. 1–6, doi: 10.1109/ICCIT60459.2023.10441093.

[39] K. Barua *et al.*, “Explainable AI-based humerus fracture detection and classification from x-ray images,” in *2023 26th International Conference on Computer and Information Technology, ICCIT 2023*, 2023, pp. 1–6, doi: 10.1109/ICCIT60459.2023.10441124.

[40] M. S. Zulfiker, N. Kabir, A. A. Biswas, T. Nazneen, and M. S. Uddin, “An in-depth analysis of machine learning approaches to predict depression,” *Current Research in Behavioral Sciences*, vol. 2, p. 100044, 2021, doi: 10.1016/j.crbeha.2021.100044.

[41] S. Hong *et al.*, “Identification of suicidality in adolescent major depressive disorder patients using sMRI: a machine learning approach.,” *Journal of Affective Disorders*, vol. 280, pp. 72–76, 2021, doi: 10.1016/j.jad.2020.10.077.

[42] A. A. Choudhury, M. R. H. Khan, N. Z. Nahim, S. R. Tulon, S. Islam, and A. Chakrabarty, “Predicting depression in bangladeshi undergraduates using machine learning,” in *Proceedings of 2019 IEEE Region 10 Symposium, TENSYMP 2019*, 2019, pp. 789–794, doi: 10.1109/TENSYMP46218.2019.8971369.

[43] A. H. Orabi, P. Buddhitha, M. H. Orabi, and D. Inkpen, “Deep learning for depression detection of Twitter users,” in *Proceedings of the fifth workshop on computational linguistics and clinical psychology: from keyboard to clinic*, 2018, pp. 88–97, doi: 10.18653/v1/w18-0609.

[44] H. Cai *et al.*, “A pervasive approach to EEG-based depression detection,” *Complexity*, vol. 2018, pp. 1–13, 2018, doi: 10.1155/2018/5238028.

[45] T. Gui *et al.*, “Cooperative multimodal approach to depression detection in Twitter,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, pp. 110–117, Jul. 2019, doi: 10.1609/aaai.v33i01.3301110.

BIOGRAPHIES OF AUTHORS

Mohammad Tarek Aziz received a B.Sc. Engineering degree from the Rangamati Science and Technology University, Rangamati-4500, Bangladesh. Currently, he is an M.Sc. Engineering student at Chittagong University of Engineering and Technology, Chattogram, Bangladesh, under the Department of Computer Science and Engineering. His research interests include ML, deep learning, digital image processing, and artificial intelligence (AI) in healthcare. He can be contacted at email: tarekaziz4288@gmail.com.

Tanjim Mahmud a member of IEEE, received both an M.Sc. and B.Sc. in Computer Science and Engineering from the University of Chittagong, Bangladesh. He is currently pursuing a Ph.D. in natural language processing at the Kitami Institute of Technology, Japan. He serves as an assistant professor in the Department of Computer Science and Engineering at Rangamati Science and Technology University, Bangladesh. His research areas encompass AI, natural language processing, AI applications in healthcare, ML, and image processing. He can be contacted at email: tanjim_cse@yahoo.com.

Md Faisal Bin Abdul Aziz holds both an M.Sc. and B.Sc. in Computer Science and Engineering from the University of Chittagong, Bangladesh. He serves as an assistant professor in the Department of Computer Science and Engineering at Comilla University, Bangladesh. His research areas encompass AI and image processing. He can be contacted at email: faisal@cou.ac.bd.

Md Abu Bakar Siddick received a B.Sc. degree from the Beijing Institute of Technology, China. Currently, he is an M.Sc. student at the same institute, under the Department of Computer Science and Technology. His research interests include data science, digital image processing, bioinformatics, and AI in healthcare. He can be contacted at email: mdsiddick6204@outlook.com.

Md. Maskat Sharif received a B.Sc. degree in Electrical and Electronics Engineering from the University of Science and Technology Chittagong (USTC), Bangladesh. Currently pursuing an M.Sc. degree in Electronics and Telecommunication Engineering under the supervision of Prof. Dr. Md. Azad Hossain. His research interests include antenna design, related circuit simulation, performance analysis, and AI. He can be contacted at email: maskat.eee@gmail.com or p22mete003@student.cuet.ac.bd.

Mohammad Shahadat Hossain a senior member of IEEE, earned his M.Phil. and Ph.D. degrees in computation from the Institute of Science and Technology at the University of Manchester (UMIST), U.K., in 1999 and 2002, respectively. He is currently a professor of Computer Science and Engineering at the University of Chittagong, Bangladesh, and serves as a visiting professor at Luleå University of Technology, Sweden. His research interests focus on e-government, risk and uncertainty modeling through evolutionary computing, the development of pragmatic software tools and methods, information systems, and expert systems. He can be contacted at email: hossain_ms@cu.ac.bd.

Karl Andersson a senior member of IEEE, obtained his M.Sc. in Computer Science and Technology from the Royal Institute of Technology in Stockholm, Sweden, and his Ph.D. in Mobile Systems from Luleå University of Technology, Sweden. Following postdoctoral work at the Internet Real-Time Laboratory, Columbia University, New York, USA, and the National Institute of Information and Communications Technology in Tokyo, Japan, he now serves as a professor at the Cybersecurity Laboratory at Luleå University of Technology. His research focuses on green and mobile computing, the internet of things, cloud technologies, and information security. He can be contacted at email: karl.andersson@ltu.se.