Bulletin of Electrical Engineering and Informatics
Vol. 15, No. 1, February 2026, pp. 895∼908
ISSN: 2302-9285, DOI: 10.11591/eei.v15i1.10010

Advanced data balancing techniques with machine learning models for acute liver failure prediction

Pradnya Borkar[1], Snehal Bankatrao Shinde[2], Mayank Jichkar[1], Mahek Humne[1], Sagarkumar Badhiye[1], Tausif Diwan[2], Nileshchandra Pikle[2]
[1]Symbiosis Institute of Technology, Nagpur Campus, Symbiosis International (Deemed University), Pune, India
[2]Department of Computer Science and Engineering, Indian Institute of Information Technology, Nagpur, India

Article Info

Article history:

Received Feb 2, 2025
Revised Oct 17, 2025
Accepted Dec 6, 2025

Keywords:

Acute liver failure
Ensemble techniques
Machine learning
Synthetic minority oversampling technique
Synthetic minority oversampling technique and edited nearest neighbours

ABSTRACT

Amongst various diseases, one of the severe diseases is acute liver failure (ALF) and it is a quick decline in liver health that normally lasts a few days to a few weeks. Machine learning (ML) techniques can play a valuable role in the diagnosis and management of ALF. The proposed study made an effort to remedy the issue of the Kaggle Dataset's class imbalance by carrying out an exhaustive experimental assessment making use of two distinct approaches, namely synthetic minority oversampling technique (SMOTE) and synthetic minority oversampling technique and edited nearest neighbours (SMOTE-ENN). Both SMOTE-balanced and SMOTE-ENN balanced datasets are used to train the support vector machine (SVM), K-nearest neighbors (KNN), logistic regression (LR), decision tree (DT), random forest (RF), eXtreme gradient boosting (XGBoost), and stacking models. Compared to the SMOTE method, the results demonstrated that the SMOTE-ENN balanced dataset achieved a considerable increase in the accuracy of its predictions. The results showed that the KNN algorithm has attained 99.52% accuracy, along with a precision of 99.07%, recall of 99.35%, and F1 measure of 99.04%. As a result, we discovered that a data balancing method that is not overly complicated and a supervised ML algorithm could be used to forecast ALF with very high accuracy and excellent potential for utility.

Corresponding Author:

Snehal Bankatrao Shinde
Department of Computer Science and Engineering, Indian Institute of Information Technology
Nagpur, Maharashtra, India
Email: ssnehalbshinde@gmail.com

1. INTRODUCTION

Immediate failure of liver is typically defined as the abrupt significant liver functions loss in a healthy individual. Numerous liver conditions, including ”advanced liver disease ,fibrosis, liver cancer and cirrhosis” can result in liver failure [1], [2]. Over 50 million individuals worldwide suffer from chronic liver diseases, making liver disease a significant health concern of our time. Hepatitis viruses are responsible for the majority of cases early stages of liver disease [3]. The main cause of liver disease is consumption of alcohol in excessive manner, being overweight, eating an unhealthy amount of red meat and fried food, and in certain instances, having a hereditary predisposition for the condition [4]-[6]. People who transition to an easy and comfortable lifestyle tend to rely heavily on unhealthy foods, which may be fatal if the liver condition is not treated soon in long run. Machine learning (ML) is essential for accurate analysis and prediction of the illness because the early symptoms are not severe and are typically neglected [7]-[9].

Large and varied datasets encompassing clinical records, lab results, imaging findings, genetic data, and biomarkers related to acute liver failure (ALF) can be analyzed using ML algorithms. ML's capacity to handle various data formats and include several variables simultaneously is one of its main advantages in ALF prediction. ML algorithms process organized and unstructured data, extract pertinent features and learn from intricate interactions between the data [10], [11]. This makes it possible to create predictive models that reflect the multidimensionality of ALF and the risk factors linked to it [12]. The accessability of diverse ML techniques and methodologies enables medical specialists to economize a significant amount of time by employing prediction from models, that assist in the assimilation of information and the production of more accurate diagnostic results [13], [14]. When determining whether or not a tumor is present, ML algorithms have fully replaced the need for radiologists, resulting in significant money savings. After extensive study and conducting in-depth examinations of previously published research, it is observed that only a few articles [15] have focused on liver failure prediction. The handling of missing values, eliminating outliers, and balancing the data using methods such as synthetic minority oversampling technique (SMOTE) and synthetic minority oversampling technique and edited nearest neighbours (SMOTE-ENN) are all data preprocessing components that are used in the proposed work. The ALF dataset usedin this study highlights the oddity of having a higher proportion of underrepresented samples in a certain class as compared to samples from other class. To produce more balanced dataset SMOTE generate psuedo data that means it creates new samplesby interpreting the samples from class minority. SMOTE-ENN used to remove misclassified samples from both the classes. The purpose of these stages is to enhance the precision of the analysis and modeling. Methods of deletion and imputation can be used to handle the issue of missing values. Extreme values, known as outliers, that can impact the findings can be identified and eliminated. To remedy the problem of class imbalance, data balancing methods such as SMOTE and SMOTE-ENN are utilized. For accurate disease classification, several well-known algorithms, such as ”random forest (RF), decision tree (DT), K-nearest neighbors (KNN), logistic regression (LR), support vector machine (SVM), and Naive Bayes (NB)” along with hyper-parameter tuning are properly trained on the preprocessed data. As a result, we recommended using conventional ML models to correctly diagnose the disease.

The motive of the research study is to apply and improve ML classification techniques to the dataset for ALF and compare the results with those of prior research that used the same dataset. The objectives of this research are summarized as follows:

a. To investigate and balance the data for analysis using SMOTE and SMOTE-ENN.
b. To compare the accuracy of the ML models KNN, LR, SVM, DT, RF, and NB by training and analyzing them on the balanced SMOTE and SMOTE-ENN dataset.
c. To improve the effectiveness of the proposed model using ensemble techniques like stacking, cat boost, RF, and eXtreme gradient boosting (XGBoost).
d. To compare the suggested methods' accuracy of prediction with the current models.
e. To develop a system that accurately predicts ALF with high precision and efficiency.

2. LITERATURE SURVEY

Research by Thirunavukkarasu *et al.* [16], different classification methods, including LR, KNN, and SVM, have been proposed. The accuracy score and the confusion matrix received the most attention. The performance is evaluated by confusion matrix and the outcome was determined based on the comparison. When compared to SVMs performance, KNN and LRs accuracy from the confusion matrix was the same, at 73.9%, while SVM had the lowest accuracy, at 71.97%. According to Adil *et al.* [17], ML model for predicting individuals suffering from liver conditions was proposed. The models included NB classifiers, DT, SVM, artificial neural network (ANN), and KNN. Additional independent variables were examined in conjunction with the statistical values that corresponded to them, such as count, mean, and standard deviation. On the test dataset consisting of 30%, performance measures were obtained for each of the algorithms. With an accuracy score of 74%, the LR performed best, while the SVM ranked worst at 58%. Research by Thaiparnit *et al.* [18] came up with a technique for classifying and utilizing the accuracy of the rule-based portion of the DT and RF ML algorithms. The technique included the one rule classification algorithm, an upgraded algorithm. This approach included a frequency table that compared each independent variable to the target variable. Each model additionally underwent a cross-validation procedure that was carried out five times. It was discovered that the RF model was the most accurate one for predicting liver dysfunction, with an accuracy of 75.76%. According to Rahman *et al.* [19], it was suggested that ML techniques like extra trees (ET), DT, and RF be used to classify liver disorders. Following this, the pearson correlation coefficient is used to select characteristics and eliminate unimportant elements from the dataset. After that, a further step called boosting, an ensemble method is done for each algorithm. Observations are made about performance indicators for these algorithms described above, including "accuracy, receiver operating characteristic-area under the curve (ROC-AUC), F1-score, precision, and recall", where the maximum accuracy was attained by LR (75%), and the lowest performance was attained by NB (53%).

In addition to classification algorithms such as LR, KNN, DT, and RF in [20] proposed "Ada-Boost, XGBoost, Light GBM, and multilayer perceptron". To conduct more accurate performance evaluations, further development of XGBoost and incorporation of a new genetic algorithm were both undertaken. The selection of additional features was carried out, and outliers were removed, which resulted in improved accuracy and brought about a reduction in the complexity of both the time and space requirements. The accuracy score of 88% was attained by the RF. Ambesange *et al.* [21] introduced the idea of utilizing hyperparameter adjustment and concentrated on the KNN supervised algorithm as their primary choice from among the many available ML algorithms. The dataset preprocessing including the feature selection is performed. Further outlier was checked, and some of it was eliminated, to ensure that the various algorithms would function smoothly. Performance metrics were calculated. Some parameters were adjusted using grid search CV, and once the model was properly trained using the parameters, an accuracy of 91% was reached. Shaheamlung *et al.* [22] suggested a model that included both supervised and unsupervised algorithms as well as reinforcement learning. They used a variety of ML techniques, including "DT, Bayesian network, ANN, J48, and SVM". The comparison that had been done before on the various datasets allowed for the drawing of a conclusion. In addition, it was found that "DT, J48, and the ANN" had all significantly increased their accuracy, with values of 98.46%, 95.04%, and 92.80%, respectively. These techniques were evaluated further with regard to their specificity and sensitivity, with the DT coming out on top with a score of 95% for specificity and ANN coming out on top with a score of 97.2% for sensitivity.

Based Zhang and Go [12], the "gradient boosting DT, XGB, and the light gradient boosting machine (LGBM)" were proposed. Each of the previously discussed boosting algorithms represented graphically, together with the hyper-parameter tweaking process, was presented. The author got to the conclusion that XGB performed best with a training time of 10.5 seconds and a maximum accuracy of 75%. In contrast, LGBM only accounted for 67% of the possible accuracy. Following the application of ML algorithms [23], [24] presented a methodology called recursive feature elimination as a means of removing features while having the least amount of damage possible. In addition to the LR, NB, SVM, RF, and multilayer perceptron ML algorithms the ensemble algorithms such as CatBoost, LGBM, XGB, and GB were proposed. The GB algorithm was the focus of much attention, and only a few features were considered. On the selected 10 features, the LGBM classifier achieves the best accuracy of 93%, while on the selected 5 features, it achieves a lower accuracy. Lastly, in [25] using 10 different ML techniques [26] using a dataset that is freely available to the public, this study classifies ALF. The dataset contains 8,785 data points. Both the training dataset and the test dataset seemed to represent superior results for the two distinct methods that we tested. After hyperparameter tuning, the RF model shows an F1-score of 99.6% for the RF approach and a training accuracy of 99.6%. The accuracy of the test, as measured by DF2, was 99.8, and the F1-score, which was calculated using the LGBM classifier, was 0.998. ALF survival prediction methods based on ML are diverse. Nevertheless, it was demonstrated that these methods had relatively limited effectiveness, with the SMOTE balancing method being the only one to produce an adequate level of effectiveness.

3. MATERIALS AND METHODS

This part includes information regarding the comprehensive dataset, exploratory data analysis, and other related methods.

3.1. Materials

This subsection provides key information regarding the dataset that was utilized in the research, including its origin, size, and properties that are pertinent to the investigation. In addition to this, it offers the pre-processing processes for the data, which involve preparing the data for future analysis. This dataset is subjected to an exploratory data analysis in order to get new insights and acquire a better understanding of its properties. It is necessary to make use of summary statistics, data visualizations, and various other relevant methods in order to conduct an exhaustive analysis of the dataset.

3.1.1. Dataset

We used a Kaggle dataset for investigation with 8,785 patient's data [27]. There are 4,155 female and 4,630 male patients in this data collection and the following characteristics describe them. People of all ages and backgrounds will benefit from a deeper comprehension of the ways that age affects the outcomes of strokes.

Age, gender, region, weight, height, body mass index, obesity, waist, maximum blood pressure, minimum blood pressure, good cholesterol, bad cholesterol, total cholesterol, dyslipidaemia, PVD: reduced circulation of blood to a body part, physical activity, education, unmarried, income, poor vision, alcohol consumption, hypertension, family hypertension, diabetes, family diabetes, hepatitis, family, hepatitis, chronic fatigue, and ALF.

3.1.2. Pre-processing of data

To prepare the data for additional analysis, these processes entail managing missing values, normalizing features, encoding categorical variables, and implementing any other.

a. Elimination of missing values: prediction quality may be negatively impacted by low-quality raw data, such as when missing values are present or when the data founds to be highly noisy. Pre-treatment procedures, such as discretizing the data, selecting relevant features, and eliminating duplicate values, are necessary to improve the suitability for data analysis. After data pre-processing, it was found that several attributes in the dataset had NULL values; as a result, it was crucial to replace any missing or null values with appropriate ones. The mean of the NULL values is used to fill in the NULL values in this case. Outliers are defined as dataset observations that considerably depart from the mean or median.
b. One hot encoding: is used to transform category variables like gender, whether a person has ever been married, the type of job they have, where they live, and if they smoke into numerical kind of data. The numerical features age, blood pressure, body mass index (BMI) can be standardized using standard scaling technique.
c. Balancing of data: the ALF dataset highlights the oddity of having a higher proportion of underrepresented samples in a certain class as compared to samples from other classes. This is the difficulty of learning a concept from a collection of instances that is little in quantity. To produce a dataset that was more balanced, the SMOTE and SMOTE-ENN techniques were applied:

– SMOTE: SMOTE generated pseudo-data that differed somewhat from the originals. It creates new samples by interpreting samples from the class minority. As a result, decision boundaries for the class minority extends into the space of the class majority, preventing over fitting. The means of class minorities typically converge to those of class majorities. The distribution and quantity of synthesized samples could alter the mean of the equalized data pool. The variability of class minority is successfully increased by SMOTE. In the equalized data pool, higher variability may result in a higher standard deviation for the class minority.
– SMOTE-ENN: the methodology can be characterised as follows for a problem involving two classes: for each sample E_i in the training set, the three nearest neighbours are found. This technique is repeated until the problem is fixed. E_i will be deleted if it belongs to the class majority and the classification assigned to it by its three closest neighbours differs from the class to which it was first assigned. If three of E_i immediate neighbours mistakenly classify E_i as belonging to a minority group, then the immediate neighbours who are part of the majority group will be excluded.

3.1.3. Exploratory data analysis

Before developing a ML algorithm, it is essential to perform exploratory data analysis to acquire a full understanding of dataset, discover any problems with the data, and locate any problems with the data.

Categorical data analysis: hepatitis and chronic fatigue in the dataset, it is observed that 27.47% of people with hepatitis and chronic fatigue have live failure and 72.53% of people are stable, whereas 4.3% of people without hepatitis and chronic fatigue have Liver Failure and 95.63% of people are not affected by it. That means the data is not balanced data. This unequal data should be balanced before applying any model to it to improve the predicting accuracy of a person with ALF who have hepatitis and chronic fatigue. Following the implementation of SMOTE, the improved results can now be seen. Bar chart Figure 1, shows that patients with ALF are 45.45% which is more than the imbalanced data. In conclusion, SMOTE used on imbalanced data, the data was balanced and the results were improved with higher accuracy of ALF prediction. In addition, to further enhance the balancing of data and to improve accuracy, SMOTE-ENN is applied. The application of SMOTE-ENN as shown in Figure 2, give that the accuracy of ALF of a person who has hepatitis and chronic fatigue is improved by 13.64% as compared to applying with SMOTE. When we use methods such as SMOTE and SMOTE-ENN to the dataset for ALF in order to balance it, we can anticipate seeing improved findings, higher accuracy in predicting ALF, and a more balanced representation of the minority class.

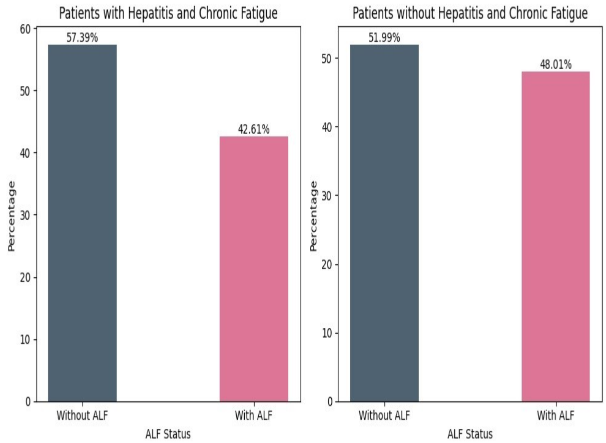

Figure 1. Distribution per hepatitis and chronic fatigue using SMOTE

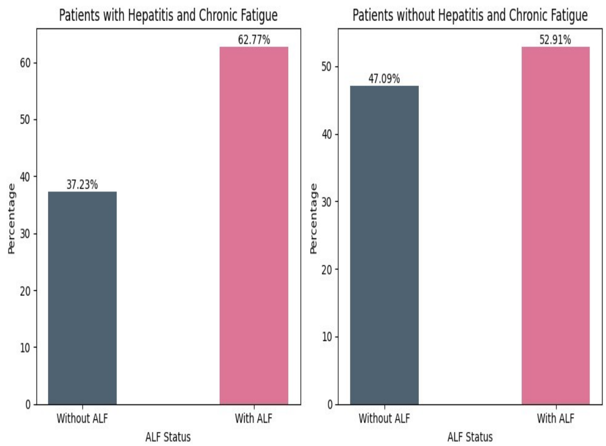

Figure 2. Distribution per hepatitis and chronic fatigue using SMOTE-ENN

3.1.4. Logistic regression

The probability of a discrete outcome using input factors is described by LR. When it comes to modeling two-valued outcomes [28] such as (true/false and yes/no), logical regression excels. Multinomial LR can be used to model situa-tions when there are more than two unique discrete outcomes. Since many issues, including threat detec-tion, may be seen as classification problems, LR is a useful analytical technique in cyber security [28]. An explanatory variable is linked to log chances in the LR model. The model of LR can be written in (1).

$$log\frac{P}{1-P} = \beta 0 + \beta 1z1 + \beta 2z2 + - - - - - + \beta nzn \,. \tag{1}$$

3.1.5. K-nearest neighbors

All of its examples are stored in memory, and it as-signs similarity scores to fresh examples when classifying them [29]. The number of nearest neighbors to be employed in the majority-rule decision is de-

termined by the KNN parameter. The distance between the two points will be determined with the use of the distance formula. It works as follows:
The formula for the distance d^2 is given as (2):

$$d^2 = (x_2 - x_1)^2 + (y_2 - y_1)^2 \quad (2)$$

The k values closest to the k factor will be selected.

3.1.6. Naïve Bayes

From Bayes' rule, NB is derived, which is a specific type of Bayesian network classifier. If the features are highly diverse, NB guarantees the maximum probability. The optimal choice rules are learned by DT through an iterative process that involves data separation based on characteristics that either maximize information gain or limit impurity.

3.1.7. Support vector machine

For quick categorization of new data points, the SVM algorithm uses some decision boundaries also called as optimal lines, which divide the n-dimensional space into distinct classes. The ideal decision boundary is hyperplane. Support vectors are often used to represent extreme cases [30].

3.1.8. Random forest

In this technique, many DT are constructed, each using subsets of the input data. The results from these trees are averaged to enhance prediction accuracy. Instead of depending on the prediction of a single DT, the RF combines the results and selects the outcome with the majority vote [31].

3.1.9. Stacking

Stacking is an aggregate method designed to integrate predictions from multiple models to enhance accuracy. This method entails training multiple base models, using their predictions to generate various features.

3.1.10. Extreme gradient boosting tree

It is a powerful ML algorithm that utilizes gradient-boosted DT. It is mostly used for supervised learning , including regression, classification, ranking. The primary goal of this algorithm is to optimize computational efficiency by leveraging the principles of gradient boosting systems. XGBoost generates DT sequentially to improve predictive accuracy [32]. In our study, we applied the HyperOpt technique for performing hyperparameter tuning for the XGB classifier [32], enhancing its performance.

3.2. CatBoost

CatBoost is a boosting algorithm specifically designed to handle datasets containing categorical variables. Most ML algorithms cannot directly process categorical or string inputs, making it necessary to convert such variables into numerical representations. CatBoost, however, has the capability to internally manage different types of categorical data [33]. It employs various statistical techniques on feature combinations to transform them into numerical values.

3.2.1. Adaptive boosting

Adaptive boosting is an algorithm mostly used as an ensemble method in ML [34]. This approach follows principle of sequential learning, where bias and variance issues in supervised learning are addressed through boosting. AdaBoost functions by merging several weak classifiers to form a single "strong" classifier. In this method, classifiers with higher accuracy are given greater weight, enhancing the overall prediction.

3.2.2. Artificial neural network

It is computational model which functions like human brain, drawing on findings from contemporary neurobiological studies. ANNs use learning and training processes for evaluation. To minimize errors, the weights of connections are adjusted iteratively [35]. This iterative process improves the accuracy of input pattern recognition, enabling reliable probability predictions.

4. PROPOSED ENHANCED ACUTE LIVER FAILURE PREDICTION MODEL WITH HYPER-PARAMETER TUNING

The suggested method consists of several steps for processing a dataset, balancing it, and applying DL and ML techniques: batch normalisation in the manner depicted in Figure 3. The following is the way to evaluate the performance of the model with the aid of execution: the first stage is data setup. The data preprocessing is done to remove outliers and fill in missing values, and then all of the attributes are converted to numerical data using one-hot encoding. Algorithm 1 illustrate how SMOTE and SMOTE-ENN approaches are used to balance the given unbalanced data. Multiple ML models including "LR, KNN, NB, DT, SVM, RF, stacking, XGBoost, CatBoost, and AdaBoost", are trained on both SMOTE-balanced and SMOTE-ENN balanced datasets with hyperparameter tuning. Additionally, training of a neural network is done using batch normalization and dropout techniques to achieve higher training performance and mitigate overfitting. Early stopping is employed to determine the optimal number of epochs.

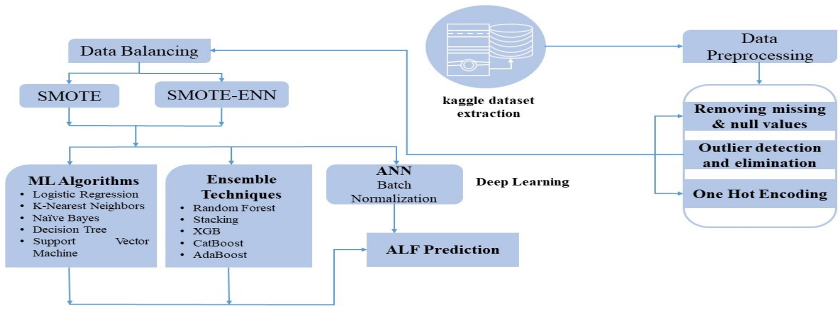

Figure 3. The structural flow of the proposed model

Algorithm 1 SMOTE-ENN algorithm for balancing and cleaning data

Input: Dataset D with majority and minority classes; Number of nearest neighbors K (default: $K = 3$)
Output: Balanced and cleaned dataset D'
Step 1: Synthetic Minority Oversampling (SMOTE)
while the dataset is imbalanced **do**
  Randomly pick an instance x from the minority class in D
  Find the K nearest neighbors of x
  for each neighbor n among the K nearest neighbors **do**
    Compute the vector difference $v = n - x$
    Draw a random scalar $r \in [0, 1]$
    Create a synthetic sample $s = x + r \cdot v$
    Add s to the dataset D
  end for
end while
Step 2: Edited Nearest Neighbor (E-NN) Cleaning
repeat
  for each instance x in D **do**
    Find the K nearest neighbors of x
    Check the majority class among these neighbors
    if the majority class is different from x's class label **then**
      Remove x from D
    end if
  end for
until the dataset achieves the desired class balance
Return D'

Various metrics, including ”accuracy, precision, recall, and F1-score”, are used to evaluate the performance of each algorithm on the testing dataset. This comprehensive evaluation aims to gauge how effectively each model performs in the context of the given problem. The ML model’s performance is assessed in the experimental setup part using Google Collaboratory [36], often known as ”Google Colab,” and Python. It provides models for segmentation, classification, prediction, and visualisation of data.

4.1. Performance metrics

The model’s correctness is assessed using a number of metrics, the majority of which are solely determined by the values of the confusion matrix. Numerous performance metrics were attained throughout the evaluation of the ML models that were considered. We shall concentrate on the terminology that appears most frequently in the relevant literature in this study [37]. Referred to as ”recall,” ”sensitivity,” or ”true positive rate,” it describes the proportion of individuals who had symptoms of acute liver and were appropriately diagnosed with the illness. A true positive (TP) result means that the anticipated class and the actual class both have value. The possibility of positive value was predicted with accuracy. TN stands for true negatives, which are a shortened form of real-class projected negative values. False positives and false negatives may result from differences between the intended class and the observed class. A case when the projected class is accurate but the actual class is erroneous is referred to as a ”false positive” (FP). When the expected class is different from the actual class, a false negative (FN) occurs.

4.2. Results

To handle the imbalanced datasets, the SMOTE and the SMOTE-ENN approaches were included in a variety of ML algorithms. When compared to just employing SMOTE, the SMOTE-ENN approach resulted in generally greater accuracy scores being reached by the majority of algorithms as shown in Table 1. KNN was able to attain the highest level of accuracy 99.2% when using SMOTE-ENN. When utilising SMOTE-ENN, RF, CatBoost, and DT all performed exceptionally well with accuracy values that were higher than 95. LR and XGBoost demonstrated large increases in accuracy with SMOTE-ENN, however, AdaBoost demonstrated a significant decrease in accuracy.

Table 1. Comparison between balancing techniques using hyperparameter tuning

Algorithms	SMOTE	SMOTE-ENN	Parameters using hyperparameter tuning
LR	92.41	94	’C’: [0.01, 0.1, 1, 10, 100], ’penalty’: [’l1’, ’l2’]
KNN	92.32	99.2	’n_neighbors’: [3,5,7], ’weights’: [’uniform’, ’distance’], ’p’: [1, 2]
NB	89.25	91.2	’var_smoothing’: [1e-9, 1e-8, 1e-7]
DT	94	96	’criterion’: [’gini’, ’entropy’], ’max_depth’: [3, 5, 7, None], ’min_samples_split’: [2, 5, 10], ’min_samples_leaf’: [1, 2, 4]
SVM	87.5	87.5	’C’: [0.1, 1, 10], ’kernel’: [’linear’, ’rbf’, ’sigmoid’], ’gamma’: [0.1, 1, ’scale’]
RF	97	98	’n_estimators’: [100, 200, 300], ’max_depth’: [None, 5, 10], ’min_samples_split’: [2, 5, 10]
Stacking	86	86.5	use_probas = True, use_features_in_secondary = True
XGBoost	94.17	97	’eta’: 0.1, ’max_depth’: 5, ’num_class’: 6, ’n_estimators’: 300, ’alpha’: 10, ’silent’: True, ’verbose_eval’: False
CatBoost	96	97.4	iterations=100, random_state=42
AdaBoost	93	87	n_estimators=50, random_state=42

5. RESULT ANALYSIS AND DISCUSSION

Comparison with the present work [25] on the dataset with our results for accuracy, precision, recall and F1-score is discussed in this section. Table 2 and Figure 4 shows the comparison of proposed algorithm's accuracy with the existing work. The accuracy of the proposed SMOTE-ENN model for LR has reached 94%, which is 13.3% higher than the accuracy that was reported for the existing work's model. For the KNN model, the proposed model has achieved high accuracy 99.2%, demonstrating that the proposed work is superior to the existing work by 4.9%. The NB model that has created and tested achieved an accuracy of 91.2, which is 15.4 points higher than the accuracy that was reported in the previous work. Both the existing work and the suggested model have achieved an accuracy of 96, which means that the accuracy of the DT is not affected by any of these models. While SVM, stacking and AdaBoost are not implemented by the existing model but proposed has shown better accuracy using SMOTE-ENN. The proposed SMOTE-ENN model increased the precision to 94%, representing an increase of 17%, from the prior work for LR, which had a precision of 77% as shown in Figure 5.

Table 2. Accuracy comparison

Algorithms	[25]	Proposed work
LR	80.7	94
KNN	94.3	99.2
NB	75.8	91.2
DT	96.8	96
SVM	NA	88
RF	99.6	98
Stacking	NA	87
XGB	84.8	97
CB	88.5	97.4
AdaBoost	NA	87

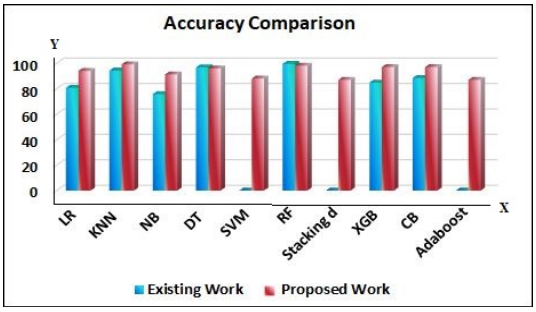

Figure 4. Accuracy comparison

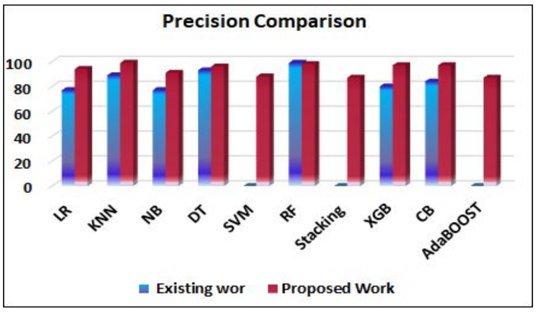

Figure 5. Precision comparison

The suggested model achieved a higher precision of 99%, exhibiting an increase of 10%, compared to the previous model's KNN precision of 89%. The new model increased precision from the prior work's 77% to 91%, which represents an increase of 14% for NB. The new model produced a slightly higher precision of 96%, exhibiting an increase of 3%, compared to the existing model's DT precision of 93%. RF has a precision of 99%, whereas the suggested model has a precision of 98%, which is roughly equivalent. XGB had attained a precision of 80% in previous studies, but the suggested model increased the precision to 97%, representing an improvement of 17%. With the previous model having a precision of 93% and CatBoost having a precision of 97%, there has been an increase of 13%. SVM, stacking, and AdaBoost are not implemented in existing work, but the suggested model has demonstrated precision that is above 85%.

The proposed SMOTE-ENN model increased the recall from the 84% achieved by LR to 94%, a 10% increase as shown in Figure 6. In the existing model, "KNN, DT, and RF" all obtained 100% recall, but the suggested model only managed a little lower recall in the range of 97-99%. In the existing model, NB had a recall of 70% however, the proposed model increased recall to 91%, a 21% improvement. Although SVM was not used in the previous study, the suggested model had an 88% recall rate. Stacking: although stacking was not used in the existing model, the suggested model had an 86% recall rate. The proposed approach increased the recall for XGBoost from 90% to 97%, which is an increase of 7%, compared to the existing work's 90% recall. CatBoost has a recall of 93%, whereas the suggested model has a recall of 97%, a 4% increase. AdaBoost was not used in the existing work, although the suggested model had an 87% recall. The proposed model scored 98%, whereas the RF work scored 99%. The SMOTE-ENN model has an F1-score of 87%, while the existing work did not apply the Stacking procedure. The SMOTE-ENN model scored 88%, while the existing work scored 85%. The model improves F1-score by 3%. In existing work, CatBoost had an F1-score of 88%, while the suggested model had 97%, a 9% gain. The proposed model has an F1-score of 87% while the existing work did not use AdaBoost as shown in Table 3 and Figure 7.

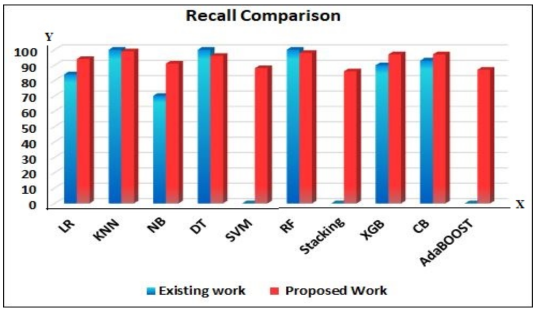

Figure 6. Recall comparison

Table 3. F1-score comparison

Algorithms	[25]	Proposed work
LR	80	94
KNN	94	99
NB	73	91
DT	96	96
SVM	NA	88
RF	99	98
Stacking	NA	87
XGB	85	88
CB	88	97
AdaBoost	NA	87

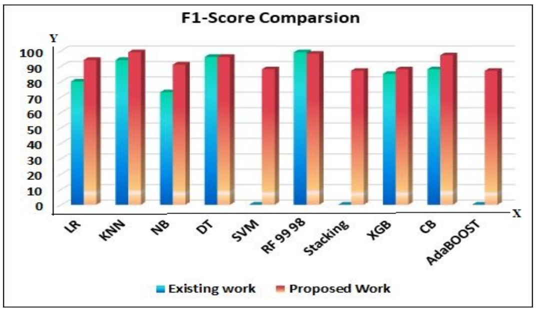

Figure 7. F1-score comparison

6. CONCLUSION

Traditional ways of diagnosing ALF depend on clinical judgment and lab tests, which may not be as accurate or reliable as possible. Large amounts of patient data, such as clinical records, medical imaging, laboratory results, and genetic information, can be analyzed by ML algorithms. It is possible to determine the most successful strategy for forecasting ALF by performing an in-depth analysis of the dataset using a number of different ML algorithms. A number of different elements have the potential to have an effect on the performance of contemporary learning systems. One of the most critical challenges is class imbalance, which occurs when instances from one class exceed those from another class in the training data by a sufficient margin. Through a comprehensive experimental review that makes use of SMOTE and SMOTE-ENN approaches, this research solves the problem of class imbalance that is present in the Kaggle dataset. For the purpose of evaluating the effectiveness of the classifiers, many measures were applied, including recall, precision, F1-score, and accuracy. These metrics are absolutely necessary for determining whether or not the models are able to effectively classify the data. The research indicates that the implementation of SMOTE and SMOTE-ENN techniques has a variety of effects on the accuracy of various ML models. These effects are outlined in the report. The results showed that the KNN algorithm has attained 99.52% accuracy, along with a precision of 99.07%, recall of 99.35%, and F1 measure of 99.04%.

According to our findings, the KNN algorithm had the highest level of accuracy. It was able to achieve this by utilizing SMOTE-ENN. In addition, the combination of SMOTE-ENN with other algorithms resulted in exceptional performance from all of the algorithms. Accuracy levels that were consistently higher than 95% were attained by RF, CatBoost, and DT. When employing SMOTE-ENN, LR, and XGBoost both displayed large improvements in accuracy; however, AdaBoost witnessed a considerable loss in accuracy. The analysis of the algorithms that are presented in the tables reveals that the suggested SMOTE-ENN model, as opposed to the SMOTE model, has consistently displayed higher performance when compared to the work that has previously been done for the majority of the metrics that have been reviewed. This would imply that the model that was proposed is more capable of handling unbalanced data and enhancing the accuracy of forecasts for ALF. This is the conclusion that can be drawn from the information presented.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Pradnya Borkar	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	
Snehal Bankatrao Shinde		✓				✓		✓	✓	✓	✓	✓		
Mayank Jichkar	✓		✓	✓		✓			✓		✓		✓	
Mahek Humne	✓		✓	✓		✓			✓		✓		✓	
Sagarkumar Badhiye	✓		✓	✓		✓			✓		✓		✓	
Tausif Diwan	✓		✓	✓		✓			✓		✓		✓	
Nileshchandra Pikle					✓		✓			✓	✓	✓	✓	✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal Analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject Administration
Fu : **Fu**nding Acquisition

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY

All data that support the findings of this study are included in the article.

REFERENCES

[1] M. G. Swain and D. E. J. Jones, “Fatigue in chronic liver disease: New insights and therapeutic approaches,” *Liver International*, vol. 39, no. 1, pp. 6–19, Jan. 2019, doi: 10.1111/liv.13919.

[2] W. M. Lee, *Acute Liver Failure*, Yamada' s Textbook of Gastroenterology, Wiley, vol. 32, pp. 1973–1988, 2015, doi: 10.1002/9781118512074.ch98.
[3] D. Castaneda, A. J. Gonzalez, M. Alomari, K. Tandon, and X. B. Zervos, "From hepatitis A to E: A critical review of viral hepatitis," *World Journal of Gastroenterology*, vol. 27, no. 16, pp. 1691–1715, 2021, doi: 10.3748/wjg.v27.i16.1691.
[4] G. Sheng, Q. Xie, R. Wang, C. Hu, M. Zhong, and Y. Zou, "Waist-to-height ratio and non-alcoholic fatty liver disease in adults," *BMC Gastroenterology*, vol. 21, no. 1, p. 239, 2021, doi: 10.1186/s12876-021-01824-3.
[5] E. J. Park *et al.*, "Dietary and Genetic Obesity Promote Liver Inflammation and Tumorigenesis by Enhancing IL-6 and TNF Expression," *Cell*, vol. 140, no. 2, pp. 197–208, 2010, doi: 10.1016/j.cell.2009.12.052.
[6] Z. Younossi *et al.*, "Global Perspectives on Nonalcoholic Fatty Liver Disease and Nonalcoholic Steatohepatitis," *Hepatology*, vol. 69, no. 6, pp. 2672–2682, 2019, doi: 10.1002/hep.30251.
[7] D. Y. Omkari and S. B. Shinde, "Opportunities and Challenges of Machine Learning and Deep Learning Techniques in Cardiovascular Disease Prediction: a Systematic Review," *Journal of Biological Systems*, vol. 31, no. 2, pp. 309–344, 2023, doi: 10.1142/S0218339023300014.
[8] A. Mittal and P. Chandra, "Improving learning in Artificial Neural Networks using better weight initializations," *International Journal of Information Technology*, vol. 17, no. 9, pp. 5519–5531, 2025, doi: 10.1007/s41870-024-01869-z.
[9] R. Cho, M. Zaman, K. T. Cho, and J. Hwang, "Investigating brain activity patterns during learning tasks through EEG and machine learning analysis," *International Journal of Information Technology*, vol. 16, no. 5, pp. 2737–2744, 2024, doi: 10.1007/s41870-024-01856-4.
[10] V. Karuppuchamy and S. Palanivelrajan, "Efficient IoT-machine learning assisted heart failure prediction using adaptive fuzzy-based LSTM-RNN algorithm," *Journal of Intelligent and Fuzzy Systems*, vol. 45, no. 1, pp. 505–520, 2023, doi: 10.3233/JIFS-224298.
[11] S. Abarna, J. I. Sheeba, and S. P. Devaneyan, "A novel ensemble model for identification and classification of cyber harassment on social media platform," *Journal of Intelligent and Fuzzy Systems*, vol. 45, no. 1, pp. 13–36, 2023, doi: 10.3233/JIFS-230346.
[12] D. Zhang and Y. Gong, "The Comparison of LightGBM and XGBoost Coupling Factor Analysis and Prediagnosis of Acute Liver Failure," *IEEE Access*, vol. 8, pp. 220990–221003, 2020, doi: 10.1109/ACCESS.2020.3042848.
[13] A. Alam and M. Muqeem, "An optimal heart disease prediction using chaos game optimization-based recurrent neural model," *International Journal of Information Technology*, vol. 16, no. 5, pp. 3359–3366, 2024, doi: 10.1007/s41870-023-01597-w.
[14] V. Behal and R. Singh, "Personalised healthcare model for monitoring and prediction of airpollution: machine learning approach," *Journal of Experimental and Theoretical Artificial Intelligence*, vol. 33, no. 3, pp. 425–449, 2021, doi: 10.1080/0952813X.2020.1744197.
[15] H. Xie, B. Wang, and Y. Hong, "A deep learning approach for acute liver failure prediction with combined fully connected and convolutional neural networks," *Technology and Health Care*, vol. 32, pp. 555–564, 2024, doi: 10.3233/THC-248048.
[16] K. Thirunavukkarasu, A. S. Singh, M. Irfan, and A. Chowdhury, "Prediction of liver disease using classification Algorithms," in *2018 4th International Conference on Computing Communication and Automation (ICCCA)*, 2018, pp. 1–3, doi: 10.1109/CCAA.2018.8777655.
[17] S. H. Adil, M. Ebrahim, K. Raza, S. S. Azhar Ali, and M. Ahmed Hashmani, "Liver Patient Classification using Logistic Regression," in *2018 4th International Conference on Computer and Information Sciences (ICCOINS)*, 2018, pp. 1–5, doi: 10.1109/ICCOINS.2018.8510581.
[18] S. Thaiparnit, N. Chumuang, and M. Ketcham, "A Comparitive Study of Clasification Liver Dysfunction with Machine Learning," in *2018 International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP)*, 2018, pp. 1–4, doi: 10.1109/iSAI-NLP.2018.8692808.
[19] A. K. M. S. Rahman, F. M. J. M. Shamrat, Z. Tasnim, J. Roy, and S. A. Hossain, "A comparative study on liver disease prediction using supervised machine learning algorithms," *International Journal of Scientific and Technology Research*, vol. 8, no. 11, pp. 419–422, 2019.
[20] M. A. Kuzhippallil, C. Joseph, and A. Kannan, "Comparative Analysis of Machine Learning Techniques for Indian Liver Disease Patients," in *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, 2020, pp. 778–782, doi: 10.1109/ICACCS48705.2020.9074368.
[21] S. Ambesange, R. Nadagoudar, R. Uppin, V. Patil, S. Patil, and S. Patil, "Liver Diseases Prediction using KNN with Hyper Parameter Tuning Techniques," in *2020 IEEE Bangalore Humanitarian Technology Conference (B-HTC)*, 2020, pp. 1–6, doi: 10.1109/B-HTC50970.2020.9297949.
[22] G. Shaheamlung, H. Kaur, and M. Kaur, "A Survey on machine learning techniques for the diagnosis of liver disease," in *2020 International Conference on Intelligent Engineering and Management (ICIEM)*, 2020, pp. 337–341.
[23] M. Eckhoff and M. Reiher, "CoRe optimizer: an all-in-one solution for machine learning," *Machine Learning: Science and Technology*, vol. 5, no. 1, pp. 1–14, 2024, doi: 10.1088/2632-2153/ad1f76.
[24] G. Shobana and K. Umamaheswari, "Prediction of Liver Disease using Gradient Boost Machine Learning Techniques with Feature Scaling," in *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, 2021, pp. 1223–1229, doi: 10.1109/ICCMC51019.2021.9418333.
[25] D. Sengupta, S. Mondal, S. Basu, A. K. De, S. Nath, and A. Pandey, "Classification of Acute Liver Failure using Machine Learning Algorithms," in *2022 IEEE Inter national Conference on Electronics, Computing and Communication Technologies (CONECCT)*, 2022, pp. 1–6, doi: 10.1109/CONECCT55679.2022.9865744.
[26] R. Fabra-Boluda, C. Ferri, M. J. Ramírez-Quintana, and F. Martínez-Plumed, "Unveiling the robustness of machine learning families," *Machine Learning: Science and Technology*, vol. 5, no. 3, pp. 1–32, 2024, doi: 10.1088/2632-2153/ad62ab.
[27] S. H. Rafi, "Acute Liver Failure," Kaggle, [Online]. Available: https://www.kaggle.com/datasets/suhailsh7/acute-liver-failure. (Accessed: Jun. 05, 2023).
[28] S. Nusinovici *et al.*, "Logistic regression was as good as machine learning for predicting major chronic diseases," *Journal of Clinical Epidemiology*, vol. 122, pp. 56–69, 2020, doi: 10.1016/j.jclinepi.2020.03.002.
[29] G. Guo, H. Wang, D. Bell, Y. Bi, and K. Greer, "KNN model-based approach in classification," in *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems*," Springer, 2003, vol. 2888, pp. 986–996, doi: 10.1007/978-3-540-39964-3_62.

[30] W. Noble, "What is a support vector machine?," *Nature Biotechnology*, vol. 24, no. 12, pp. 1565–1567, 2006.
[31] Y. Qi, "Random Forest for Bioinformatics," *Ensemble Machine Learning*, pp. 307–323, 2012, doi: 10.1007/978-1-4419-9326-7_11.
[32] T. Chen and T. He, "XGBoost: Extreme Gradient Boosting," *R Lecture version 0.4-2*, vol. 1, no. 4, pp. 1–4, 2015.
[33] A. V. Dorogush, V. Ershov, and A. Gulin, "CatBoost: gradient boosting with categorical features support," *arXiv preprint*, 2018, doi: 10.48550/arXiv.1810.11363.
[34] Y. Cao, Q.-G. Miao, J.-C. Liu, and L. Gao, "Advance and Prospects of AdaBoost Algorithm," *Acta Automatica Sinica*, vol. 39, no. 6, pp. 745–758, 2014, doi: 10.3724/sp.j.1004.2013.00745.
[35] A. K. Jain, J. Mao, and K. M. Mohiuddin, "Artificial neural networks: A tutorial," *Computer*, vol. 29, no. 3, pp. 31–44, 1996, doi: 10.1109/2.485891.
[36] E. Bisong, "Building Machine Learning and Deep Learning Models on Google Cloud Platform," Berkeley, CA: Apress, 2019, pp. 59–64, doi: 10.1007/978-1-4842-4470-8.
[37] M. Hossain, "A review of literature on open innovation in small and medium-sized enterprises," *Journal of Global Entrepreneurship Research*, vol. 5, no. 1, pp. 34–50, 2015, doi: 10.1186/s40497-015-0022-y.

BIOGRAPHIES OF AUTHORS

Dr. Pradnya Borkar is a prominent academic with a Ph.D. in Computer Science and Engineering from RTMNU, Nagpur focusing on bioinformatics. Her research area focuses on parallel computing, high performance computing and bioinformatics with numerous publications in esteemed and conferences. She is also a co-inventor of a patented system for an artificial intelligence enabled robot for home assistance. Also she has co-authored one book and edited one book. She can be contacted at email: pradnyaborkar2@gmail.com.

Dr. Snehal Bankatrao Shinde is a distinguished academic with a Ph.D. in Computer Science and Engineering from VNIT, Nagpur, focusing on systems biology. She has taught at notable institutions such as Vishwakarma University and Vellore Institute of Technology. Her research spans ML and deep learning in healthcare, with numerous publications in esteemed journals and conferences. She is also a co-inventor of a patented system for ECG classification, child safety, brain stroke prediction, and smart-hard disk, sepsis prediction using deep learning. Currently, she is working as an Assistant Professor at IIIT Nagpur. She can be contacted at email: sshinde@iiitn.ac.in.

Mayank Jichkar B.Tech. 4th year student in Computer Science with a keen interest in ML and its applications. Contributed to impactful research, including a project on ALF prediction. This work, which leverages ML algorithms like logistic regression and DT, is also part of a research paper aimed at enhancing early diagnosis and intervention in critical healthcare scenarios. Expertise lies in applying AI to solve real-world problems, particularly in the healthcare domain. He can be contacted at email: mayankjichkar1105@gmail.com.

Mahek Humne a final-year B.Tech. student in Computer Science with a strong passion for ML and its innovative applications. She has contributed to significant research efforts, including a project focused on ALF prediction. This project, utilizing ML techniques like logistic regression and DT, forms part of a research paper aimed at advancing early detection and timely intervention in critical healthcare situations. Her expertise centres on harnessing AI to address real-world challenges, especially in the healthcare sector. She can be contacted at email: humnemahek@gmail.com.

Sagarkumar Badhiye is a dedicated researcher with expertise in artificial intelligence, data science, automata, and diabetic retinopathy. His work emphasizes applying advanced computational techniques to solve real-world problems, particularly in healthcare and data-driven systems. He has authored several research papers in reputed journals and conferences, contributing to the advancement of these domains. He is also actively involved in interdisciplinary research, focusing on bridging gaps between academia and industry to foster innovation and practical applications of emerging technologies. He can be contacted at email: sagarbadhiye@gmail.com.

Tausif Diwan received his M.Tech. in Computer Science and Ph.D. in Parallel Computing from Department of Computer Science and Engineering, Visvesvaraya National Institute of Technology, Nagpur, India in 2011 and 2017 respectively. Since July 2019, he has been associated with the Department of Computer Science and Engineering, Indian Institute of Information Technology, Nagpur as an Assistant Professor and currently working as head of the department. His research areas include parallel computing, ML, and deep learning. He has published several research papers in various international conferences and reputed journals. An Australian Patent titled "An IoT-based Smart Food-Cold Chain Transportation System" is also granted to his account. He has associations and collaborations with several industries. He is NVidia certified DLI instructor for the "Foundation of deep learning" and "Building Transformer-Based Natural Language Processing Applications" courses and he secured "Platinum" award as a DLI Instructor from Nvidia in the year 2022. He can be contacted at email: tdiwan@iiitn.ac.in.

Nileshchandra Pikle is a distinguished figure in the field of computer science, renowned for his expertise in deep learning and GPU computing. Holding a Ph.D. in Computer Science and Engineering. He specializes in GPU programming, with a particular focus on the parallelization of finite element method (FEM) on CUDA-enabled GPUs, a subject on which he has published papers perception and cognition, and neurobiology of learning and memory. He can be contacted at email: nilesh.pikle@gmail.com.