2022 IEEE Smartworld, Ubiquitous Intelligence & Computing, Scalable Computing & Communications, Digital Twin, Privacy Computing, Metaverse, Autonomous & Trusted Vehicles (SmartWorld/UIC/ScalCom/DigitalTwin/PriComp/Metaverse)

Learning High-dimensional Associations for Nonalcoholic Fatty Liver Disease Diagnosis Prediction

1st Zhijin Wang*
College of Computer Engineering
Jimei University
Xiamen, China
zhijinecnu@gmail.com

2nd Bing Cai
College of Computer Engineering
Jimei University
Xiamen, China
ccailearning@gmail.com

3rd Wen Yang
Xiamen Hospital of Traditional Chinese Medicine
Fujian University of Traditional Chinese Medicine
Fuzhou, China
ywen99012@gmail.com

4th Peisong Zhang
School of Science
Jimei University
Xiamen, China
pencil007123@gmail.com

5th Yaohui Huang
College of Electronic Information
Guangxi Minzu University
Nanning, China
yhhuang5212@gmail.com

6th Jinmo Tang
Xiamen Hospital of Traditional Chinese Medicine
Fujian University of Traditional Chinese Medicine
Xiamen, China
king.m.t@163.com

***Abstract*—Nonalcoholic Fatty Liver Disease (NAFLD) is a major liver disease worldwide, and NAFLD diagnosis is the clinical foundation before healthcare strategies. NAFLD diagnosis predictive methods provide a promising way for intelligent diagnosis. Previous diagnoses mainly relied on liver biopsy and imaging, which are unsuitable for large-scale rapid screening. However, the observed quantitative data are high-dimensional and have missing values. To alleviate this issue, the stacking multi-scaled convolutional neural network (SMCNN) is proposed to predict NAFLD diagnosis. Firstly, the inputs are normalized and transformed. Secondly, the inputs are represented as multi-scaled feature maps. Finally, the scaled feature maps are stacked to connect with outputs. Several experiments have been done to validate twelve methods on a real dataset. The proposed SMCNN outperforms eleven methods in terms of five metrics.**

***Index Terms*—Deep learning, High dimensionality, NAFLD, Prediction, Smart diagnosis**

I. Introduction

Nonalcoholic Fatty Liver Disease (NAFLD) is the most common liver disease, which affects approximately one-quarter of the population worldwide [1]. NAFLD encompasses a wide range of disease spectrum, which includes nonalcoholic fatty liver (NAFL), nonalcoholic steatohepatitis (NASH), cirrhosis, and hepatocellular carcinoma [2]. The rising NAFLD prevalence increases the healthcare and economic burdens for human beings.

Early diagnosis and treatment can greatly alleviate the situation. The current clinical NAFLD diagnosis mainly relies on liver biopsy and imaging techniques, such as ultrasound (US), transient elastography (TE), computed tomography (CT), magnetic resonance imaging (MRI) and their developments [3]. Recently, machine learning techniques have been quickly developed and combined with the conventional diagnosis methods [4]. For example, Convolutional neural network (CNN) and its variations [5] have been widely applied in the identification of pathological sections and ultrasound images.

Liver biopsy is traumatic and expensive for screening [6]. The imaging devices are costly and not suitable for large-scale screening studies. Moreover, NAFLD is usually diagnosed when patients suffer from the more severe disease since the NAFLD has no obvious symptoms. Hence, a promising way is to leverage universal and quantitative data for diagnosis prediction, such as blood tests, urine tests and laboratory data. Several studies have been proposed to utilize these data to predict NAFLD diagnosis, such as NASH-Scope [7].

The problem of NAFLD diagnosis prediction can be commonly considered as a binary classification problem. Several machine learning methods [8] have been proposed to predict NAFLD diagnosis by leveraging routine examined data. However, there are three challenges:

1) *High-dimensionality*. The inputs consist of hundreds of clinical blood-examined items.
2) *Data missing*. Patients usually do not check all items, which leads to the problem of missing data.
3) *Data uncertainty*. Different patients are sensitive or insensitive to different items. The personalized characters of patients degrade the prediction performance.

To address these issues, the stacked multi-scaled convolutional neural network (SMCNN) is proposed to predict NAFLD diagnoses. Firstly, the clinical examined data is

This work was supported in part by the Natural Science Foundation of Fujian Province (CN) (nos. 2020J05146, 2021J01857, 2021J01859 and 2022J01335) and the Natural Science Foundation of China (no. 62006096).

979-8-3503-4655-8/22/$31.00 ©2022 IEEE

DOI 10.1109/SmartWorld-UIC-ATC-ScalCom-DigitalTwin-PriComp-Metaverse56740.2022.00052

extracted from clinical records. The inputs are normalized. Secondly, a feature selection layer is employed to select the strongly correlated data. Thirdly, a stacked multi-scaled convolutional layer is utilized to represent the inputs into multi-scaled feature maps to learn the associations between high-dimensional features. Finally, the convolutional representations are fed into full-connection and activation layers to connect with outputs. Extensive experiments are conducted to evaluate the effectiveness of the proposed SMCNN on real NAFLD diagnosis dataset.

The major contributions of this paper are listed as follows:

1) A promising way for large-scale NAFLD screening is tested.
2) The stacked multi-scaled convolutional neural network (SMCNN) is proposed to predict NAFLD diagnoses with better performance.
3) Some experiments have been done to validate the prediction method.

The rest of this paper is organized as follows. Section II reviews relevant studies on NAFLD diagnosis. Section III displays the problem definition, method workflow and describes the key details. Section IV gives the experimental configurations. Section V shows the experimental results and analyses. Finally, the conclusions are drawn in Section VI.

II. Related Work

The relevant studies are introduced with respect to the clinical NAFLD diagnosis methods and the research on the NAFLD diagnosis prediction.

A. Clinical NAFLD Diagnosis

The clinical NAFLD diagnosis methods can be categorized into two classes according to the invasiveness.

The invasive method refers to the liver biopsy, which is the gold standard for diagnosing liver diseases [9]. The doctors diagnose the NAFLD via combining observations of pathological sections and clinical records. However, liver biopsy has several limitations, such as invasiveness, sampling error, only moderate intraobserver, and interobserver reproducibility [10]. Therefore, non-invasive methods like imaging have rapidly developed and replaced liver biopsy in some specific cases. Non-invasive methods mainly rely on two different approaches the biological approach and the physical approach. The biological methods diagnose NAFLD according to the quantification of biomarkers in serum samples [11]. These methods generate NAFLD diagnosis by combining the biomarkers into mathematical algorithms, such as fatty liver index (FLI), NAFLD fibrosis score, and BARD score [12]. The accuracy is the major limitation of these methods. Conventional imaging methods, like US, CT, and MRI, have been well-established. Moreover, these methods have several limitations, such as subjectivity and examiner-dependent interpretation [13]. In the past decade, updated imaging methods have the ability to quantify liver diseases using both ultrasonography and MRI-based methods, such as ultrasound elastography, quantitative ultrasound techniques, magnetic resonance elastography, and magnetic resonance-based fat quantitation techniques [14]. These methods combine the imaging and quantitative index to diagnose NAFLD, which provides a more objective and interpretable way for medical intervention. (NAFLD fibrosis score, Fib-4 score, BARD score, and others)

However, all of these methods are not suitable for large-scale screening since these methods are still relied on the examiner's interpretation, huge time consuming, and expensive cost.

B. Computer-aid NAFLD Diagnosis

The current computer-aid methods on the NAFLD diagnosis prediction can be categorized into image-based and quantitative methods.

The image-based methods mainly focus on employing computer vision techniques to recognize the disease signatures from pathological section or liver imaging for diagnosis prediction [15]. For instance, CNN and its variations [5] have been widely employed in extracting and identifying the histological features from pathological section images. The major limitation of these methods in practical scenarios is the invasiveness of the patients. CNN [5] have also been employed to extract disease signatures from ultrasound images. These methods all are the combinations of clinical diagnosis and machine learning techniques, which show high accuracy in NAFLD diagnosis prediction. However, these methods are not suitable for large-scale screening for NAFLD.

The quantitative methods aim to build a mapping between quantitative inputs and the diagnosis target via representations [16]. Traditional machine learning methods obtain the diagnosis predictions according to some feature representations by utilizing quantitative data, such as Random forest (RF), Support vector machine (SVM), and XGBoost (XGB) [17]. These methods have simple logic and low complexity but are sensitive to the missing values. Since the ability to capture non-linear relationships, Artificial neural network (ANN) [7] have been employed to diagnose the NAFLD. These methods represented the routine laboratory or physical examination data into high-dimensional vector representations to generate diagnosis predictions according to the learned feature associations. These methods provide a potential way for mass screening.

However, they did not consider the effects of data imbalance and sparsity caused by the high dimensionality and data uncertainty.

III. The Proposed SMCNN

A. Problem Formulation

The NAFLD diagnosis is the process that determines whether the outpatients have it or not. Therefore it can be commonly regarded as a binary classification problem. Let

symbol $\boldsymbol{X} \in \mathbb{R}^{N\times M}$ represent the cleaned metrics matrix, where N is the number of inpatients, and M is the number of cleaned metrics. The binary classification can be formulated as follows:

$$\boldsymbol{Y} \leftarrow F(\boldsymbol{X}), \tag{1}$$

where $\boldsymbol{Y} \in \mathbb{R}^{N}$ is the label matrix, $F(\cdot)$ is a learned mapping.

The scheme illustrations of the proposed method are shown in Figure 1. In the proposed method, the diagnosis is first fed into a data clean layer to initially filter out invalid and missing values. Secondly, the cleaned data is passed through the normalization layer to reduce the difference between metrics values. And then, the normalized matrix is fed into a multi-convolutional layer to extract feature representations. Finally, the representations are passed through a full connection layer and activation layer to generate classification results, the major symbols and notations used in this paper are listed in Table I.

TABLE I
SYMBOLS AND SEMANTICS.

Notation	Meaning
N	Samples number
M	Examination-items number
K	Filtered examination-items number
$\boldsymbol{X}$	Input matrix, where $\boldsymbol{X} \in \mathbb{R}^{N\times M}$
$\boldsymbol{Z}$	Normalized input matrix, $\boldsymbol{Z} \in \mathbb{R}^{N\times M}$
$\boldsymbol{U}$	Filtered normalized input matrix $\boldsymbol{U} \in \mathbb{R}^{N\times K}$
H	Feature maps number
$\mathcal{G}(\cdot)$	Pooling strategies
$f(\cdot)$	Flatten operation
$\boldsymbol{P}$	Probability matrix, $\boldsymbol{P} \in \mathbb{R}^{N\times 1}$
$\boldsymbol{Y}$	Binary target matrix, $\boldsymbol{Y} \in \mathbb{R}^{N\times 1}$

B. Normalization

To reduce the difference between the metrics and speed up the convergence process, the "Min-Max normalization" [18], [19] is adopted to compress the inputs into $[0, 1]$. The mathematical formulation is defined as follows:

$$\boldsymbol{Z}_i = \frac{\boldsymbol{X}_i - \min(\boldsymbol{X}_i)}{\max(\boldsymbol{X}_i) - \min(\boldsymbol{X}_i)}, \tag{2}$$

where $\boldsymbol{Z}_i$ represents the i-th normalized item, $\boldsymbol{X}_i$ is the the i-th input item, $\min(\boldsymbol{X}_i)$ is the minim value of $\boldsymbol{X}_i$, $\max(\boldsymbol{X}_i)$ is the maxim value of $\boldsymbol{X}_i$.

C. Feature Selection

To alleviate the effect of the high dimensionality and improve the model accuracy, feature selection techniques are adopted to transform the input matrix to be small-scale. The filter methods are employed for feature selection by ordering, which leverages variable ranking techniques as the criterion. The chi-square test is chosen as the correlation criterion. The mathematical definition is as follows:

$$\mathcal{X}^2 = \frac{(\boldsymbol{Z}_i \cdot \boldsymbol{Y} - \frac{1}{N}\sum_{j=1}^{N} Y^j \cdot \sum_{j=1}^{N} Z_i^j)^2}{\frac{1}{N}\sum_{j=1}^{N} Y^j \cdot \sum_{j=1}^{N} Z_i^j}, \tag{3}$$

where $\boldsymbol{Z}_i$ is the i-th feature matrix, $\mathcal{X}^2$ is the correlation score matrix. And then the features are selected accord to magnitude of the $\mathcal{X}^2$. Let symbol $\boldsymbol{U} \in \mathbb{R}^{K}$ be the filtered feature matrix, where K is the number of top K ranked features, $\boldsymbol{Z} \in \mathbb{R}^{N\times M}$ is the normalized input matrix.

D. Stacked multi-scaled convolutional Unit

To reduce the effect of sparse matrix and build associations between inputs. A stacked multi-scaled convolutional unit (SMCU) is designed. The detailed process of the multi-scaled convolutional layer is plotted in Figure 1(b). The proposed SMCU is mainly constituted of convolutional layers and pooling layers.

For a lucid presentation, given the input matrix $\boldsymbol{x}$, the i-th feature map sweeps through the input matrix $\boldsymbol{x}$ can be formulated as follows:

$$\mathcal{C}_i^r(\boldsymbol{x}) = ReLU(\boldsymbol{W}_i^r * \boldsymbol{x} + b_i), \tag{4}$$

where $W_i^r \in \mathbb{R}^{1\times r}$ is the kernel weight, $b_i \in \mathbb{R}$ is the bias term of i-th feature map, $*$ denotes the convolution operation, the $ReLU(\cdot)$ is the rectified linear unit function and $\mathcal{C}_i^r(\cdot)$ refers to the i-th feature map with kernel size r.

And the pooling layer with pooling kernel size k is defined as follows:

$$\mathcal{R}_i^k(\boldsymbol{x}) = \mathcal{G}(\boldsymbol{x}_{i,j-1}^k, \boldsymbol{x}_{i,j}^k, \cdots, \boldsymbol{x}_{i,j+k-1}^k), \tag{5}$$

where $\boldsymbol{x}$ represents the output of preceding layer, $\mathcal{G}(\cdot)$ represents the pooling strategy, p represent the pooling size, $\mathcal{R}_i^k(\cdot)$ refers to pooling i-th feature map with size k.

The *residual learning* trick is adopted to alleviate the degradation problem. The graphical process of this process is plotted in Figure 1(c). The residual unit is formulated as follows:

$$\mathcal{E}_i^r(\boldsymbol{x}) = \mathcal{C}_i^r(\boldsymbol{x}) + ReLU(\boldsymbol{x}), \tag{6}$$

where $\boldsymbol{x}$ represents the output of preceding layer, $\mathcal{E}_i^r$ refers to the residual unit to the i-th feature map.

For the proposed SMCU, the filtered input matrix $\boldsymbol{U}$ is firstly passed through a convolutional layer and max-pooling layer to initially explore the association between inputs. The process can be formulated as follows:

$$\boldsymbol{S}_i = \mathcal{R}_i^2(C_i^7(\boldsymbol{U})), \tag{7}$$

where $\boldsymbol{S}_i$ is the i-th feature map outputs, $\boldsymbol{S} \in \mathbb{R}^{N\times H\times K|2}$ is the output matrix, H is the number of feature maps.

And then, the feature maps are passed through two convolutional layer blocks with smaller kernel sizes to learn the associations between features further. After being well-tuned, the kernel size is determined at 5 and 3, respectively. To ensure stability and performance, the residual learning trick is employed. The process can be formulated as follows:

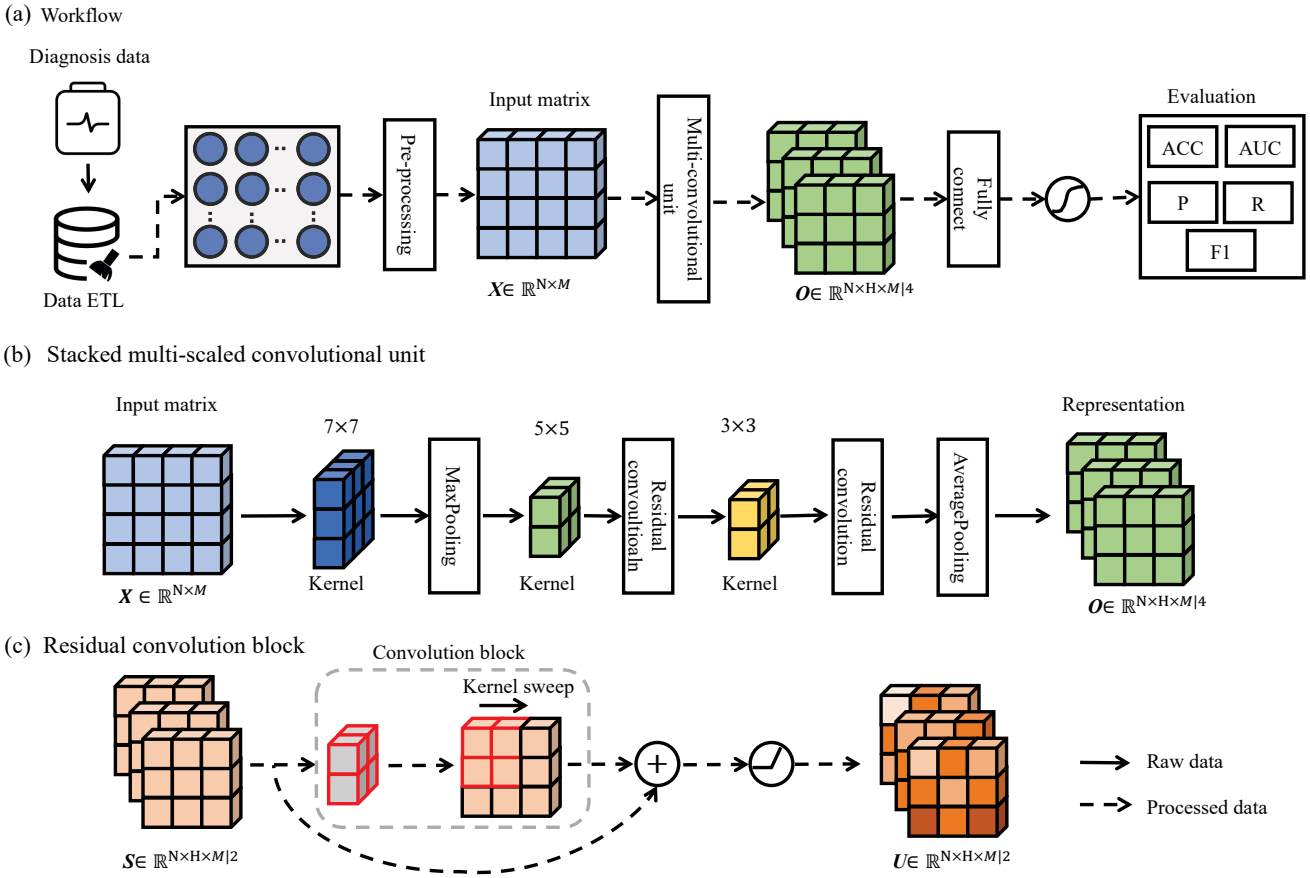


Fig. 1. The schematic illustration of the proposed method. (a) The workflow. (b) Stacked multi-scaled convolutional unit (SMCU) (c) Residual convolution block.

$$\boldsymbol{I}_i = \mathcal{E}_i^3(\mathcal{E}_i^5(\boldsymbol{S})), \tag{8}$$

where $\boldsymbol{I} \in \mathbb{R}^{N\times H\times K|2}$ is the proceeding block outputs.

Finally, the feature maps $\boldsymbol{I}$ are fed into an average-pooling layer to learn the association between feature representations further. The process can be formulated as follows:

$$\boldsymbol{A}_i = \mathcal{R}_i^2(\boldsymbol{I}), \tag{9}$$

where $\boldsymbol{A} \in \mathbb{R}^{N\times H\times K|4}$ represents the pooled feature maps, $\mathcal{R}_i^2(\cdot)$ represents the pooling operation.

E. Full connection layer

To aggregate feature maps and inputs, the *Full-connection* is selected as the feature fusion method since it is easy and robust. The process can be formulated as follows:

$$\boldsymbol{O} = f([\boldsymbol{A}_1, \boldsymbol{A}_2, \cdots, \boldsymbol{A}_H]) \odot \boldsymbol{W}_p + B_p, \tag{10}$$

where $\boldsymbol{O} \in \mathbb{R}^{N\times 1}$ is the prediction matrix, $\boldsymbol{W}_p \in \mathbb{R}^{H\cdot M|4\times 1}$ is the weight matrix, $f(\cdot)$ is the flatten operation, B_p is the bias term.

The outputs of the proceeding layer are fed into a sigmoid layer to generate probability matrix $\boldsymbol{P}$ and generate classification results. These processes are formulated as follows:

$$P_i = \sigma(O_i) = \frac{1}{1+\exp^{-O_i}}, \tag{11}$$

$$Y_i = \begin{cases} 0, & P_i < 0.5 \\ 1, & P_i \geq 0.5 \end{cases}, \tag{12}$$

where $\boldsymbol{Y} \in \mathbb{R}^{N\times 1}$ is the classification results.

F. Loss Function

The Binary Cross-Entropy with Logits Loss (BCEwith-LogitsLoss) is chosen as the loss function for the content criterion. The BCEwithLogitsLoss is constituted of a Binary Cross-Entropy Loss function and a Sigmoid function. The mathematical formulation of the loss function is as follows:

$$\boldsymbol{L}(\boldsymbol{O}, \boldsymbol{Y}) = \frac{1}{n}\{-\boldsymbol{w}[\boldsymbol{Y} log\sigma(\boldsymbol{O}) + (1-\boldsymbol{Y})log(1-\sigma(\boldsymbol{O}))]\}, \tag{13}$$

where $\boldsymbol{L}$ is the loss matrix, $\top$ represents the transpose operation, $\boldsymbol{w}$ is the weight matrix, σ is the Sigmoid function.

IV. Experimental Configurations

A. Dataset

NAFLD diagnosis is shared by Xiamen Hospital of Traditional Chinese Medicine (XHTCM). In the 304 real-world outpatients' electronic health records, a total of 338 physical examination items were collected to evaluate the proposed method and benchmarks. These items include a blood routine examination, blood biochemistry examination, liver function test, renal function test, blood glucose test, blood lipid test, electrolytes, potassium, sodium, chloride, myocardial enzymes, and urine examination. Notably, the required items are different for outpatient. Therefore, these data are usually high-dimensional with missing values. For some quantitative items, the values are directly collected, such as white blood cell count (WBC) and red blood cell count (RBC). For the binary text description, positiveness and negativeness are replaced with 1 and 0, respectively, such as white blood cells (WBC) and red blood cells (RBC).

We collected data from March 29, 2018, to April 28, 2022. All individual-level data are anonymized. The diagnostic record dataset is homogeneously divided into five parts for the cross-validation strategy [20]. One of them was selected to validate the trained models.

B. Benchmarks

The below machine learning methods are adopted as parts of benchmarks.

1) *K-nearest neghibors (KNN)* [21] is a non-parametric methods, which generate classifications according to the distances between features.
2) *SVM* [22] maps the feature into high-dimensionality and classifies the samples according to the distances between samples and hyperplane.
3) *RF* [23] uses partial features to train multiple decision trees and generate classification results by tree voting.
4) *Gradient boosting decision tree (GBDT)* [24] is an iterative decision tree algorithm, which generates classifications according to multiple decision trees.
5) *XGB* [25] minimizes the loss function by iteratively fitting the annotations and the residuals.
6) *Gaussian naive bayes (GNB)*, *Multinomial naive bayes (MNB)* and *Bernoulli naive bayes (BNB)* are all Bayes-based methods [26] and the difference between them is the likelihood estimation of features.
7) *AdaBoost (Ada)* [27] makes a series of weak learners a strong learner by repeatedly modifying the weights of the data.

The below deep learning methods are adopted as parts of benchmarks.

7) *ANN* [7] maps the features through the multi-layer hidden neural to minimizes the loss function.
8) *Transformer (TRFM)* [28] is a variations of encoder-decoder structure, which captures key features via attention mechanism.

C. Evaluation Metrics

The Accuracy (ACC), Precision (P), Recall (R), F1-Means (F1) and Area under the ROC Curve (AUC) values are chosen to evaluate the performance of the binary classification. The higher values of these metrics are, the better performance of the methods. The mathematical representations of evaluation metrics are listed in the Table II and Table III. For a lucid presentation, the mathematical formulation of AUC is indecently listed as follows:

$$AUC = \frac{2 \cdot S_p - n_p \cdot (n_n + 1)}{2 \cdot n_p \cdot n_n}, \tag{14}$$

where S_p is the sum of the all positive examples ranked, n_p and n_n represent the number of positive and negative examples respectively.

TABLE II
Confusion Matrix for Binary Classification

Real \ Predicted	False	True
False	FP	TN
True	FN	TP

TABLE III
Evaluation Metrics of Classification Performance for the Proposed Method and Benchmarks.

Metrics	Formula
ACC	$\frac{TP+TN}{TP+TN+FP+FN}$
Precision	$\frac{TP}{TP+FP}$
Recall	$\frac{TP}{TP+TN}$
F1-Means	$\frac{2 \cdot P \cdot R}{P+R}$
TPR	$\frac{TP}{TP+FN}$
FPR	$\frac{FP}{FP+TN}$

V. Experimental Results

A. Feature Selection

To explore the effective feature and provide a fountain for further clinical research, the performance is measured in terms of AUC, ACC, and $F1$ by varying feature number K. The collaboration of coarse-tune and fine-tune strategy is adopted to find the optimal K. The experiments results are visualized in Figure 2. As shown in Figure 2(f)-2(e), the optimal values of K is found at 69. And the correspond items can be detected according to the strengthen of the correlation coefficient. Several observations from the experiment are summarized as follows:

1) The blood test and urine test data have strong correlations with with the NAFLD incident. Specifically, the urine pH value has the strongest correlations with the NAFLD.
2) Several liver diseases metrics have strong correlated with the NAFLD incident.

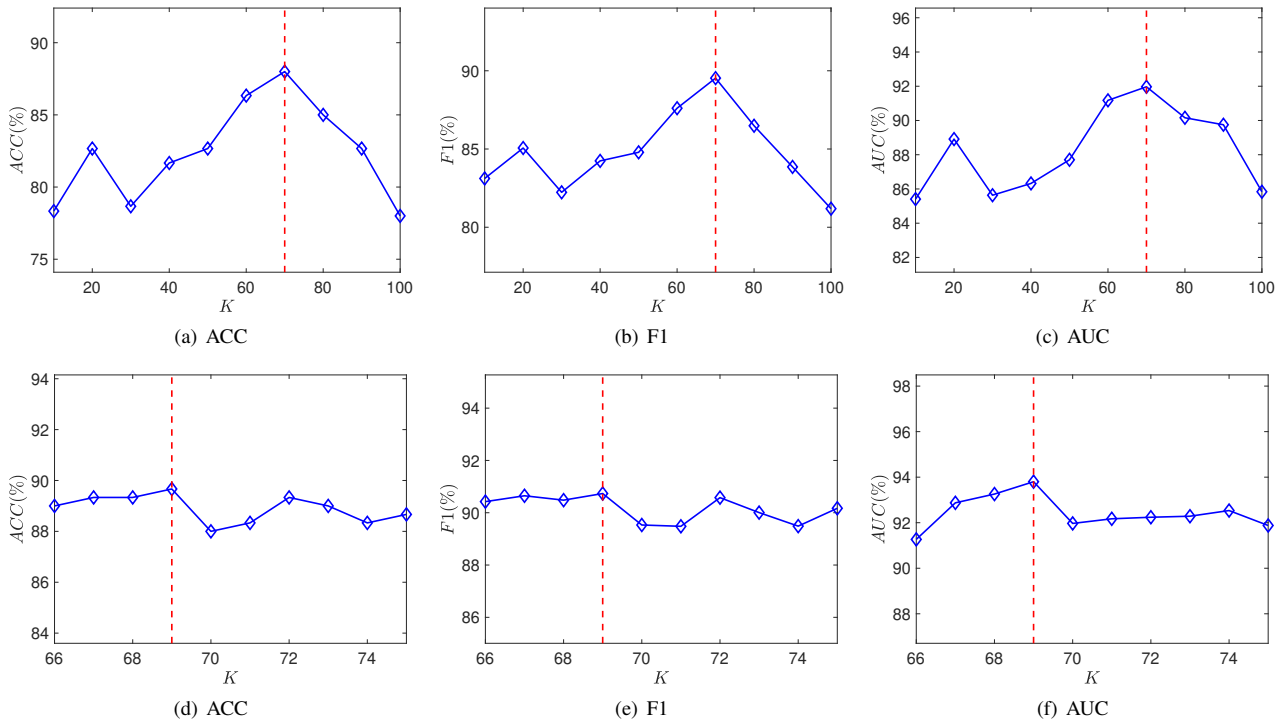


Fig. 2. The performance of the proposed SMCNN with varying the feature number K in terms of AUC, ACC, and $F1$.

3) Numerous liver-uncorrelated metrics are beneficial for improving the NAFLD diagnosis prediction accuracy.

The routine examination metrics like red blood cell count (RBC), white blood cells (WBC), urea, and the other blood test and urine test metrics show strong correlations with the NAFLD. A possible reason is that the NAFLD would affect the metabolism, which causes these metrics to be abnormal. The urine pH value strongly correlates with the NAFLD. A possible reason is that this metric has associations with insulin resistance, which is associated with NAFLD [29]. What's more, the blood glucose (Glu) and urea/creatinine show strong correlations with the NAFLD in the experiment. The above metrics all have associations with diabetes. And some institutes [30] pointed out that diabetes patients have a higher proportion of having the NAFLD. In summary, the collected samples possibly have a large proportion of people with diabetes.

Several liver disease-related metrics are beneficial for improving the diagnosis accuracy, such as γ-glutamyl transpeptidase (γ-GT), hepatitis B virus surface antibody (HBsAb), and albumin (Alb). There are several possible reasons:

1) The NAFLD is usually detected when the patients search medical intervention for some more serious liver disease, such as hepatitis, cirrhosis.
2) The hepatitis may increase the risk of the NAFLD.
3) The NAFLD could cause serious liver diseases, such as cirrhosis, hepatitis. With the development of the NAFLD, these metric start to be abnormal.

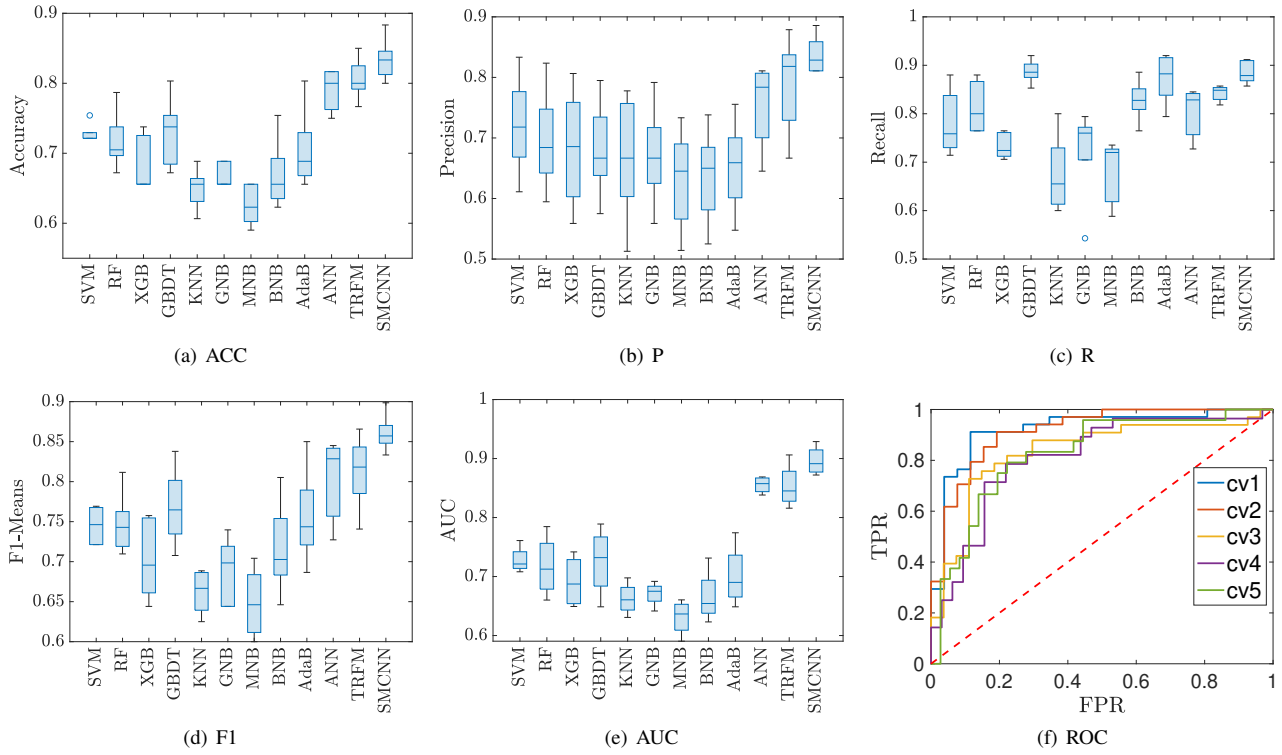
Therefore, the metrics for these diseases are highly correlating to the NAFLD.

Several liver-uncorrelated disease metrics, such as typhoid H, typhoid O, and mycoplasma IgG, seem effective for diagnosis prediction. There are two possible reasons, and one is that these diseases bring indirect effects for the NAFLD. For example, typhoid fever would significantly affect the gut microbiota, while the gut microbiota is closely related to the development of the NAFLD [31]. The details should be further researched and explored in clinical. Another is that the patients suffering from these diseases are diagnosed with NAFLD. And the data scale is not large enough, which causes these diseases to be associated with NAFLD.

B. Comparable Results

The performance of comparison is done to explore the effectiveness of the proposed SMCNN. The comparison results of twelve methods are plotted in Figure 3(a)-3(e). For all comparable methods, all parameters are well-tuned. Besides, feature-filtered inputs are used to train all methods. Several observations from the experiment results are summarized as follows:

1) The Bayes-related methods achieves the worst performance among benchmarks, especially the GNB.
2) The boosting-based methods obtain the highest R value.
3) Deep learning methods are more stable and efficient than machine learning methods in general.



(a) ACC (b) P (c) R

(d) F1 (e) AUC (f) ROC

Fig. 3. Box-plots of eleven methods in terms of Accuracy, Precision, Recall, F1-Means, and Area under curve values are plotted in 3(a)-3(e). The ROC lines of the proposed method is plotted in 3(f), where cv_i represent the i-th validation subset.

4) The proposed SMCNN obtained significant improvements compared with benchmarks.

According to Figure 3, the Bayes-related methods achieve the worst performance. This demonstrates that the joint probability distribution of inputted features has difference from a specific probability distribution, such as Gaussian or Bernoulli. A possible reason is the effect of missing values, and zero values would interrupt the probability distribution.

The boosting-related methods outperform the other benchmarks, especially the performance evaluated by R. However, the other metrics' values can not match the R. This means that these methods can accurately recognize the positive samples but cannot recognize the negative samples well.

As shown in Figure 3(a)-3(e), the deep learning methods obtain a smaller interquartile range compared with machine learning methods. This demonstrates that deep learning methods are more robust and stable. Moreover, the medium performance of the deep learning methods is higher than the major machine learning methods in terms of P, F1, and AUC values. This shows the effectiveness of the deep learning methods and demonstrates that deep learning methods do better in capturing effective features from high-dimensional data and alleviating the data sparsity.

The proposed SMCNN achieves measurable improvements compared with benchmarks. The ACC, P, and F1 values are improved by 3.3%, 6.31%, and 5.97% compared with the optimal benchmarks. This demonstrates that the proposed multi-scaled convolutional structure does better in extracting features from sparse and high-dimensional data.

To further explore the performance of the proposed classifier, the area under curve (AUC) is adopted to evaluate all methods, and the receiver operating characteristic (ROC) curve of test subsets are plotted in Figure 3(e) and 3(f). As shown in Figure 3(e), the proposed SMCNN achieved the highest median AUC value, and the AUC value is improved by 4.33% compared with the optimal benchmarks. This suggests that the proposed method has an excellent ability in classification. Moreover, as shown in Figure 3(f), the curves all are located in the above red dash line. These further demonstrate the effectiveness of the proposed method. In addition, the distribution of curves is similar, which shows the stability of the SMCNN.

VI. Conclusion

This paper proposed a multi-scaled convolutional neural network (SMCNN) for predicting the NAFLD diagnosis by solely using physical examination data. The proposed multi-scaled convolutional layer can efficiently learn the associations between high dimensionality and alleviate the effect of data sparsity. Extensive experiments on real NAFLD examination-items dataset are done to explore the effectiveness of the proposed method. Compared with the benchmarks, the ACC, P, $F1$, and AUC values is improved by

3.3%, 6.31%, 5.97% and 4.33% at least, respectively. Moreover, experimental analysis shows several liver-uncorrelated metrics have strong correlations with the NAFLD.

In the future, the feature selection and feature fusion methods will be further discussed in NAFLD diagnosis prediction, and the multi-diseases diagnosis prediction also would be further studied.

ACKNOWLEDGMENT

Thanks to the Xiamen Hospital of Traditional Chinese Medicine (XMTCM) for sharing the data. We gratefully appreciate the editor and anonymous reviewers for their valuable insights and suggestions which enormously benefited this paper.

REFERENCES

[1] T. G. Cotter and M. Rinella, "Nonalcoholic fatty liver disease 2020: The state of the disease," *Gastroenterology*, vol. 158, no. 7, pp. 1851 – 1864, 2020.

[2] Q. M. Anstee, H. L. Reeves, E. Kotsiliti, O. Govaere, and M. Heikenwalder, "From nash to hcc: current concepts and future challenges," *Nature reviews Gastroenterology & hepatology*, vol. 16, no. 7, pp. 411 – 428, 2019.

[3] R. Loomba, "Role of imaging-based biomarkers in nafld: recent advances in clinical application and future research directions," *Journal of hepatology*, vol. 68, no. 2, pp. 296–304, 2018.

[4] A. M. Dinani, K. V. Kowdley, and M. Noureddin, "Application of artificial intelligence for diagnosis and risk stratification in nafld and nash: The state of the art," *Hepatology Research*, vol. 74, no. 4, pp. 2233–2240, 2021.

[5] T. Jiang, X. jing Guo, L. ping Tu, Z. Lu, J. Cui, X. xiang Ma, X. juan Hu, X. hua Yao, L. tao Cui, Y. zhi Li, J. bin Huang, and J. tuo Xu, "Application of computer tongue image analysis technology in the diagnosis of nafld," *Computers in Biology and Medicine*, vol. 135, p. 104622, 2021.

[6] N. F. Schwenzer, F. Springer, C. Schraml, N. Stefan, J. Machann, and F. Schick, "Non-invasive assessment and quantification of liver steatosis by ultrasound, computed tomography and magnetic resonance," *Journal of Hepatology*, vol. 51, no. 3, pp. 433 – 445, 2009.

[7] T. Okanoue, T. Shima, Y. Mitsumoto, A. Umemura, K. Yamaguchi, Y. Itoh, M. Yoneda, A. Nakajima, E. Mizukoshi, S. Kaneko *et al.*, "Artificial intelligence/neural network system for the screening of nonalcoholic fatty liver disease and nonalcoholic steatohepatitis," *Hepatology Research*, vol. 51, no. 5, pp. 554 – 569, 2021.

[8] L. Yifang, W. Xuetao, Z. Jun, Z. Shanshan, and J. Jian, "Applications of artificial intelligence (ai) in researches on non-alcoholic fatty liver disease(nafld) : A systematic review," *Reviews in Endocrine and Metabolic Disorders*, vol. 23, pp. 387 – 400, 2022.

[9] Y. Zhou, J. Zheng, W. Liu, L. Miele, A. Vitale, S. Van Poucke, T.-T. Zou, D.-H. Fang, S. Shen, D.-C. Zhang, and M.-H. Zheng, "The nafl risk score: A simple scoring model to predict 4-y risk for non-alcoholic fatty liver," *Clinica Chimica Acta*, vol. 468, pp. 17–24, 2017.

[10] B. A. Davison, S. A. Harrison, G. Cotter, N. Alkhouri, A. Sanyal, C. Edwards, J. R. Colca, J. Iwashita, G. G. Koch, and H. C. Dittrich, "Suboptimal reliability of liver biopsy evaluation has implications for randomized clinical trials," *Journal of hepatology*, vol. 73, no. 6, pp. 1322–1332, 2020.

[11] V. Kupčová, M. Fedelešová, J. Bulas, P. Kozmonová, and L. Turecký, "Overview of the pathogenesis, genetic, and non-invasive clinical, biochemical, and scoring methods in the assessment of nafld," *International Journal of Environmental Research and Public Health*, vol. 16, no. 19, 2019.

[12] S. Wenjing, C. Hongli, L. Ning, W. Yanling, L. Shujie, Y. Yang, Y. Xinru, and C. Dong-Feng, "Comparison of fib-4 index, nafld fibrosis score and bard score for prediction of advanced fibrosis in adult patients with non-alcoholic fatty liver disease: A meta-analysis study," *Hepatology Research*, vol. 46, no. 9, pp. 862–870, 2016.

[13] T. Nobuharu, A. Veeral, and L. Rohit, "Non-invasive methods for imaging hepatic steatosis and their clinical importance in nafld," *Nature Reviews Endocrinology*, vol. 18, pp. 55–66, 2022.

[14] L. Qian, D. Manish, G. Joseph, R, S. Claude, and S. Anthony, E, "Current status of imaging in nonalcoholic fatty liver disease," *World Journal of Gastroenterology*, vol. 10, no. 8, pp. 530 – 542, 2018.

[15] E. M. Brunt, A. D. Clouston, Z. Goodman, C. Guy, D. E. Kleiner, C. Lackner, D. G. Tiniakos, A. Wee, M. Yeh, W. Q. Leow, E. Chng, Y. Ren, G. G. Boon Bee, E. E. Powell, M. Rinella, A. J. Sanyal, B. Neuschwander-Tetri, Z. Younossi, M. Charlton, V. Ratziu, S. A. Harrison, D. Tai, and Q. M. Anstee, "Complexity of ballooned hepatocyte feature recognition: Defining a training atlas for artificial intelligence-based imaging in nafld," *Journal of Hepatology*, pp. 1 – 12, 2022.

[16] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[17] G. L.-H. Wong, P.-C. Yuen, A. J. Ma, A. W.-H. Chan, H. H.-W. Leung, and V. W.-S. Wong, "Artificial intelligence in prediction of non-alcoholic fatty liver disease and fibrosis," *Journal of Gastroenterology and Hepatology*, vol. 36, no. 3, pp. 543–550, 2021.

[18] P. Zhang, Z. Wang, Y. Huang, and M. Wang, "Dual-grained directional representation for infectious disease case prediction," *Knowledge-Based Systems*, vol. 256, p. 109806, 2022.

[19] Z. Wang, Y. Huang, and B. He, "Dual-grained representation for hand, foot, and mouth disease prediction within public health cyber-physical systems," *Software: Practice and Experience*, vol. 51, no. 11, pp. 2290 – 2305, Oct 2021.

[20] Z. Wang, Z. Wang, L. Lu, Y. Huang, and Y. Fu, "A multi-view multi-omics model for cancer drug response prediction," *Applied Intelligence*, vol. Early View, 2022.

[21] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, "Scikit-learn: Machine learning in python," *the Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011.

[22] J. Cervantes, F. García-Lamont, L. Rodríguez-Mazahua, and A. López Chau, "A comprehensive survey on support vector machine classification: Applications, challenges and trends," *Neurocomputing*, vol. 408, pp. 189–215, 2020.

[23] J. L. Speiser, M. E. Miller, J. Tooze, and E. H. Ip, "A comparison of random forest variable selection methods for classification prediction modeling," *Expert Syst. Appl.*, vol. 134, pp. 93–101, 2019.

[24] H. Rao, X. Shi, A. K. Rodrigue, J. Feng, Y. Xia, M. Elhoseny, X. Yuan, and L. Gu, "Feature selection based on artificial bee colony and gradient boosting decision tree," *Appl. Soft Comput.*, vol. 74, pp. 634–642, 2019.

[25] L. Munkhdalai, T. Munkhdalai, and K. H. Ryu, "Gev-nn: A deep neural network architecture for class imbalance problem in binary classification," *Knowledge-Based Systems*, vol. 194, p. 105534, 2020.

[26] H. Dou, J. Tan, H. Wei, F. Wang, J. Yang, X.-G. Ma, J. Wang, and T. Zhou, "Transfer inhibitory potency prediction to binary classification: A model only needs a small training set," *Computer Methods and Programs in Biomedicine*, vol. 215, p. 106633, 2022.

[27] L. Shihua, D. Zian, C. Tianxin, C. Hongyi, and J. Ling, "A weighted svm ensemble predictor based on adaboost for blast furnace iron-making process," *Applied Intelligence*, vol. 50, no. 7, pp. 1997–2008, 2020.

[28] N. Mapes, A. White, R. Medury, and S. Dua, "Divisive language and propaganda detection using multi-head attention transformers with deep learning BERT-based language models for binary classification," in *EMNLP'19*, Hong Kong, China, Nov 2019, pp. 103–106.

[29] T. Okamura, Y. Hashimoto, M. Hamaguchi, A. Obora, T. Kojima, and M. Fukui, "Low urine ph is a risk for non-alcoholic fatty liver disease: A population-based longitudinal study," *Clinics and Research in Hepatology and Gastroenterology*, vol. 42, no. 6, pp. 570–576, 2018.

[30] D. o. H. Union Medical College Hospital and Medical, "Health talk - fatty liver," https://www.pumch.cn/detail/13635.html, 2017, (accessed 2017).

[31] A. Pérez-Montes de Oca, M. T. Julián, A. Ramos, M. Puig-Domingo, and N. Alonso, "Microbiota, fiber, and nafld: Is there any connection?" *Nutrients*, vol. 12, no. 10, 2020.