IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 7 April 2023, accepted 19 April 2023, date of publication 27 April 2023, date of current version 4 May 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3271137

RESEARCH ARTICLE

A Novel Approach for Cardiotocography Paper Digitization and Classification for Abnormality Detection

SIBEL ÖZTÜRK[1], SAFIYE AĞAPINAR ŞAHIN[1], AYŞE NUR AKSOY[2], BERNA ARI[3], AND ALEX AKINBI[4]

[1]Department of Midwifery, Atatürk University, 25030 Erzurum, Turkey
[2]Department of Obstetrics and Gynecology, Erzurum Regional Training and Research Hospital, 25240 Erzurum, Turkey
[3]Department of Electrical and Electronics Engineering, Firat University, 23119 Elâzığ, Turkey
[4]School of Computer Science and Mathematics, Liverpool John Moores University, L2 2QP Liverpool, U.K.

Corresponding author: Alex Akinbi (o.a.akinbi@ljmu.ac.uk)

This work was supported in part by the Ataturk University and Erzurum City Hospital Turkiye Participatory Research Project under Grant TKP-2022-10965.

ABSTRACT Cardiotocography (CTG) is a clinical procedure that is used to track and gauge the severity of fetal distress. Although CTG is the most often used equipment to monitor and assess the health of the fetus, the high rate of false positive results due to visual interpretation significantly contributes to needless surgical delivery or delayed intervention. In this study, a novel approach is introduced where both printing CTG paper is digitized and a machine learning approach is employed to detect the abnormality in the digitized CTG signal. Image processing-based preprocessing steps are employed to make the printing of CTG paper more convenient to extract the CTG signal. Various signal-processing techniques are used to calibrate the extracted CTG signal. Then, Empirical Mode Decomposition (EMD) is used to decompose the CTG signal into its frequency components and instantaneous frequency and spectral entropy features are extracted. After feature normalization and feature selection with ReliefF algorithm, support vector machines (SVM) is used for the classification of the normal and abnormal classes. A novel dataset is used in the experimental works and various performance evaluation metrics are used for the evaluation of the achievement of the proposed method. 10-fold cross-validation-based experiments show that the proposed method is quite efficient in abnormality detection in printing CTG papers where an average accuracy score of around 90.0% is produced.

INDEX TERMS EMD, feature selection, image enhancement, printing CTG paper, signal reconstruction, SVM classifier.

I. INTRODUCTION

Cardiotocography (CTG) is a widely used technique for monitoring fetal well-being during pregnancy and labor. The primary aim of CTG is to assess fetal heart rate (FHR) and uterine contractions (UC) to detect potential complications and ensure the optimal health of the fetus [1], [2]. The traditional method of CTG involves the use of paper strips to record FHR and UC, which are manually interpreted by healthcare professionals [3]. This approach, however, has several limitations, including inter- and intra-observer variability in interpretation, lack of standardization, and limited ability to detect subtle changes in fetal well-being. The development of signal digitalization of CTG paper has emerged as a promising solution to overcome these challenges. Digitalization of CTG paper involves converting analog signals recorded on paper into digital format, which can be analyzed by computer algorithms [4].

The resulting digital signal can be visualized on a computer screen and analyzed using various quantitative parameters. This approach offers several benefits, including increased accuracy and reproducibility, standardized interpretation, and the ability to detect subtle changes in fetal well-being.

The associate editor coordinating the review of this manuscript and approving it for publication was Kumaradevan Punithakumar.

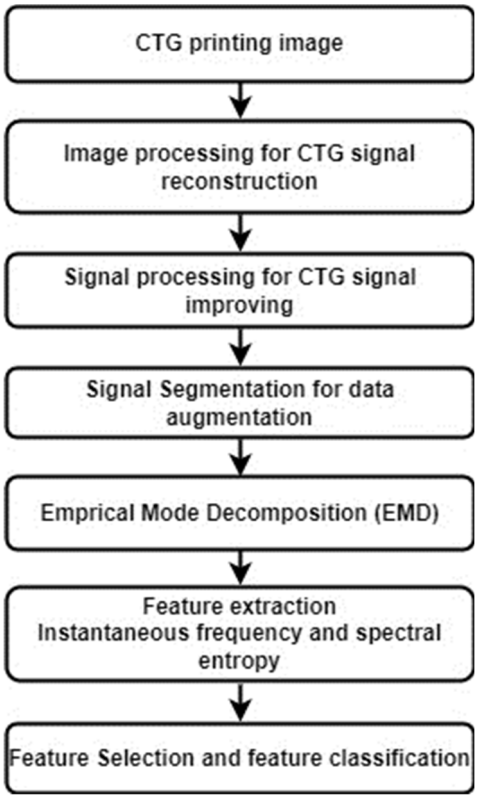


FIGURE 1. Block diagram of the proposed method.

The objective of this study is to develop an accurate and robust artificial intelligence-based system to read, evaluate and classify CTG print images as normal or abnormal. Thus, a novel approach is developed for both CTG paper digitization and classification of the digitized CTG signal into normal and abnormal classes, respectively. Fig. 1 shows the block diagram of the proposed method. While the CTG paper digitization stage is based on image processing, the classification of the CTG signals is rely on machine learning. The input to the developed system is the scanned CTG paper and the output is its class label. It is worth mentioning that only the FHR signal is used in our developed system, so manual cropping is carried out to acquire the FHR signal part of the scanned CTG paper. The RGB color image space is initially converted to Lab color space as the Lab color space is designed to be perceptually uniform, meaning that a change of the same amount in any direction in the color space will result in a perceived color difference of approximately the same magnitude.

This makes it easier to make adjustments to the color and brightness of an image more intuitively. Then, median filtering is applied to each Lab color channel separately. And L channel is weighted (multiplied by a constant value) to increase the lightness of the input image. Then the Lab color space is converted to RGB color space again. After these preprocessing operations, only the red channel is used to extract the signal from the red grid background. Some morphological operations are used to remove some unwanted binary noise and the signal amplitude calibration is carried out to determine the amplitude of the signal in beats per minute (BPM) scale. After signal extraction, an interpolation operation is employed to fill the missing points of the signal accordingly [15].

And all signals are resampled to fix their lengths to a constant duration. A multiresolution signal analysis approach namely Empirical Mode Decomposition (EMD) is employed to the extracted FHR signals to decompose them into their frequency components and for each decomposed signal two feature extraction methods namely instantaneous frequency and spectral entropy are used to extract features from the normal and abnormal FHR signals. A feature selection procedure is employed for determining the most efficient feature set. The well-known ReliefF approach is used for feature selection. Finally, the SVM classification approach is used in the detection of abnormalities. Accuracy, sensitivity, specificity, and F1-score metrics are used for the performance evaluation of the proposed method. The main contributions of this study are;

1-) An accurate Machine Learning (ML) system is proposed for CTG printing image-based fetus abnormality detection.
2-) Windowing-based data augmentation and multi-resolution-based approach are used for increasing the efficiency of CTG-based abnormality detection.

In the next section, the related works, the proposed method and the related theories will be introduced. The experimental works and results are given in Section IV. Discussions and conclusions are given in Sections V and VI, respectively.

II. RELATED WORKS

In recent years, several studies have investigated the potential benefits of signal digitalization of CTG paper in clinical practice.

One study conducted by Verburg et al. [5] compared the performance of digitalized CTG with traditional CTG in detecting fetal distress. The study found that digitalized CTG was significantly more accurate than traditional CTG in detecting fetal distress, with a sensitivity of 94.9% compared to 73.0% for traditional CTG.

Austin et al. [6] used machine learning algorithms to analyze digitalized CTG traces to predict adverse neonatal outcomes. The study found that machine learning algorithms were able to accurately predict adverse neonatal outcomes, including low Apgar scores and neonatal intensive care unit admission. These results suggest that signal digitalization of CTG paper has the potential to improve fetal outcomes by enabling earlier detection of fetal distress and appropriate interventions.

Cifuentes et al. [7] used artificial intelligence (AI) algorithm to analyze digitalized CTG traces to predict fetal acidemia. The study found that the AI algorithms had a high accuracy in predicting fetal acidemia, with a sensitivity of 87.5% and specificity of 86.2%. These findings suggest that

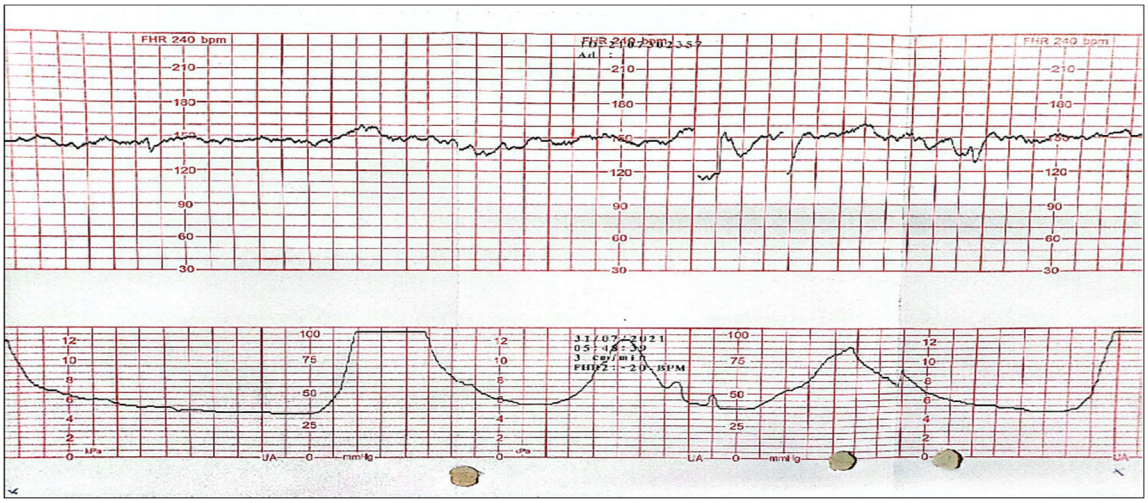


FIGURE 2. A sample scanned CTG printing paper.

AI algorithms could be a valuable tool in clinical practice for predicting adverse fetal outcomes.

Ayres-de-Campos et al. [8] developed a signal-processing algorithm to analyze digitalized CTG traces and predict fetal acidemia. The study found that the algorithm had a sensitivity of 88.2% and a specificity of 85.9% in predicting fetal acidemia. These results suggest that signal processing algorithms could be a valuable tool for predicting adverse fetal outcomes.

Ge et al. [9] used a deep learning model to classify CTG signals as normal or abnormal based on FHR and UC patterns. The model achieved a classification accuracy of 92.15%, demonstrating the potential of deep learning techniques for CTG interpretation.

Rana et al. [10] used an ensemble of machine learning classifiers to classify CTG signals as normal, suspicious, or pathological. The ensemble achieved a classification accuracy of 98.4%, indicating the potential of machine learning methods for automated CTG interpretation.

Chudáček et al. [11] developed an algorithm to classify CTG signals as normal, suspicious, or pathological based on FHR and UC patterns. The algorithm achieved a classification accuracy of 87.3%, indicating its potential for clinical use.

Another study by Chudáček et al. [12] developed an algorithm to classify CTG signals as normal, suspicious, or pathological based on FHR variability and acceleration. The algorithm achieved a classification accuracy of 88.8%, demonstrating the potential of more specific CTG signal analysis for improving classification accuracy.

Amer-Wahlin et al. [13] found that automated CTG interpretation had a higher false-positive rate compared to visual interpretation by experienced clinicians. Therefore, the development of automated CTG interpretation systems should be accompanied by rigorous validation and testing to ensure their accuracy and reliability in clinical practice. The hybrid approach for digitizing CTG signals was proposed by Comert et al. [14]. The proposed method comprises two basic steps, one based on image processing and the other on signal processing. In the image processing stage, image filtering and image segmentation techniques were utilized, while signal calibration was used in the signal processing step. The average correlation coefficients for the FHR and UC signals were 0.9715±0.0168 and 0.9717±0.0465, respectively, according to the authors' analysis of 552 CTG recordings.

From the reviewed literature, it was seen that the accuracies of the proposed methods were stacked mostly within the 80% and 90% bands. So, more accurate methods are required for producing reliable CTG-based abnormality detection systems for early warning of both patients and clinicians.

III. RESEARCH METHODOLOGY

As mentioned earlier, in this study, a two-staged approach was proposed for abnormality detection in CTG printing papers. These stages were signal extraction from the printed CTG papers and classification of the extracted signal into normal and abnormal classes, respectively. The input to the signal extraction stage was the printed CTG papers, more specifically, the FHR part [16]. And the input to the signal classification part was the extracted FHR signal. As the first part contains various image processing algorithms, the second part was based on machine learning (feature extraction and feature classification). Fig. 2 shows a sample of printed CTG paper that was used as an input to the first part. As seen in Fig. 1, the CTG printing papers contain both FHR (up) and UC (down) signals.

While the FHR signal was scaled on the 30-240 BPM, the UC signal was scaled on 0-100 mmHg. Fig. 3 shows the output image after applying the image-based preprocessing routines. As observed from Fig. 3, the colors were now more enhanced and the background noise was eliminated. The signal trace was also more distinguishable from the background grid. As the FHR signal was used for abnormality detection,

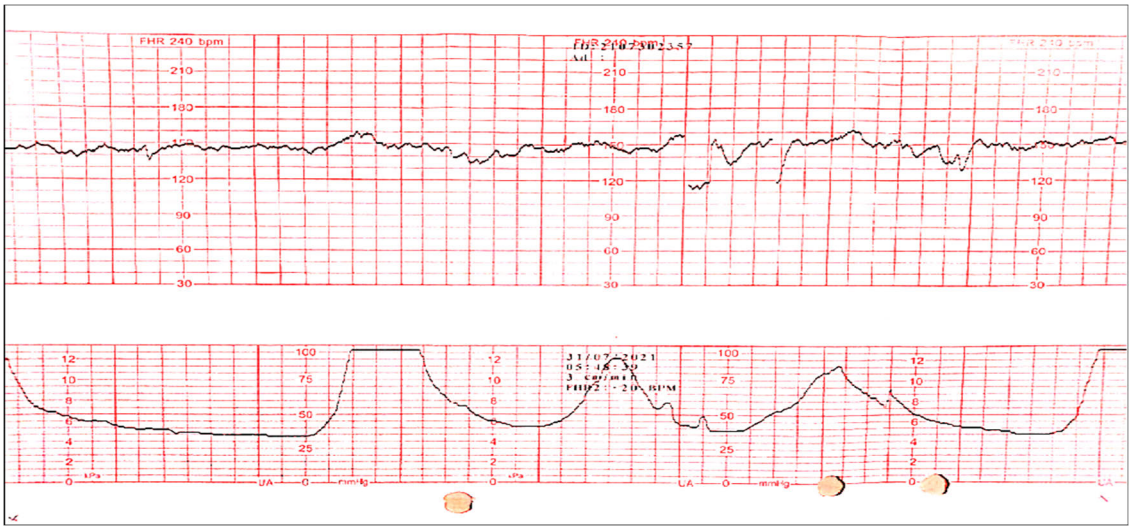


FIGURE 3. The preprocessed input printing CTG paper.

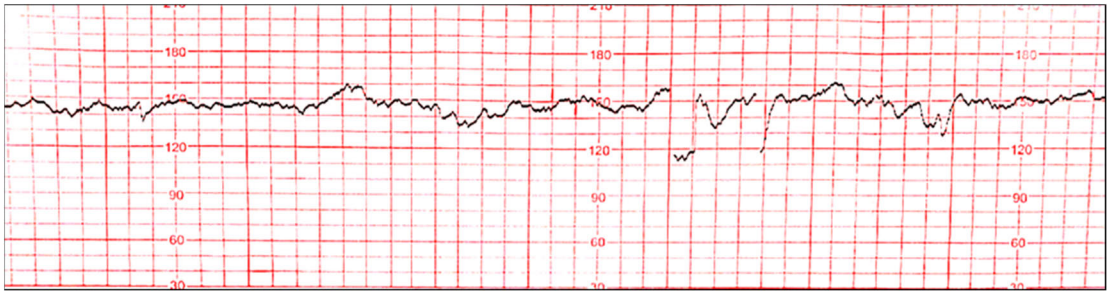


FIGURE 4. The cropped FHR region from the printing CTG paper.

manual cropping was employed to acquire the FHR region from the CTG printing paper. Fig. 4 shows the cropped FHR region from the CTG printing paper. As seen in Fig. 4, the 30-240 BPM scale was cropped and further used in abnormality detection.

A. PREPROCESSING

The input printing CTG paper image was initially converted to the Lab color space and the L channel is weighted and the other channels (a and b) are preserved. By doing so, the lightness of the input image was enhanced. Then each L, a, and b channel are filtered separately by using a symmetric 2D median filtering where the window size is selected as 5×5. After filtering, the Lab color space was reconverted to the Red, Green, and Blue (RGB) color space for further processes. In RGB color space, a contrast enhancement operation was applied to the R, G, and B channels. While the red channel contrast was limited to the 0.15 and 0.22 range, the green and blue channels' contrasts are limited to the 0-1 range.

B. LAB COLOR SPACE

Lab color space was a color model used in digital image processing and computer graphics. It is a device-independent color model, which means that it was not tied to any specific device, such as a particular camera, printer, or display. Lab stands for Luminance, a, and b, which are the three components that make up the color space. The Luminance (L) component represents the brightness of the color, with values ranging from 0 to 100, where 0 is black and 100 is white. The a and b components represent the color channels, with a ranging from green to red and b ranging from blue to yellow. The Lab color space is often used in color management systems to help maintain consistent color reproduction across different devices. It allows for a wider gamut of colors to be represented than other color models such as RGB or CMYK and is particularly useful for color correction and adjustment in digital images.

C. OTSU THRESHOLDING

After preprocessing, the red channel of the enhanced FHR image was considered for further processes. The red channel is considered because the printing CTG paper has a red grid background and the red channel only contains the signal in a white background as shown in Fig. 5. And a simple thresholding method can be used to extract the signal from the background. To this end, the well-known Otsu thresholding is used [17], [18].

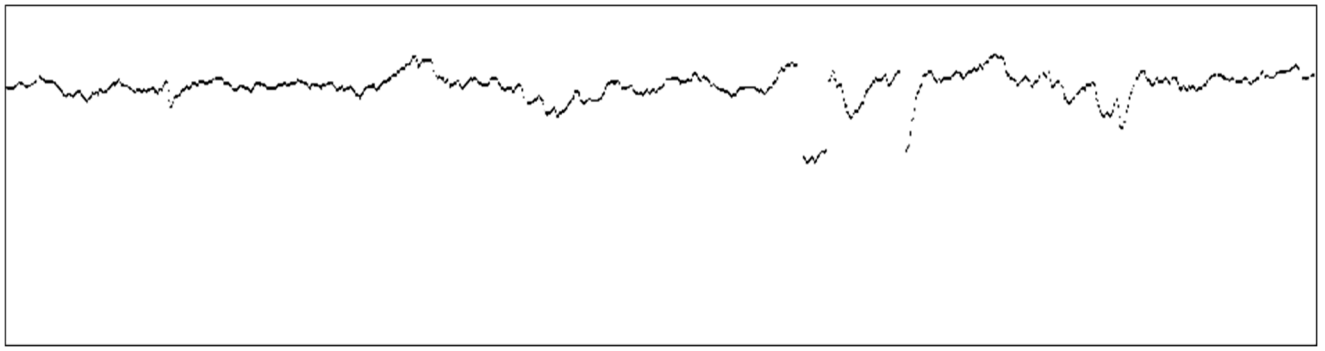


FIGURE 5. The red channel of the preprocessed FHR part.

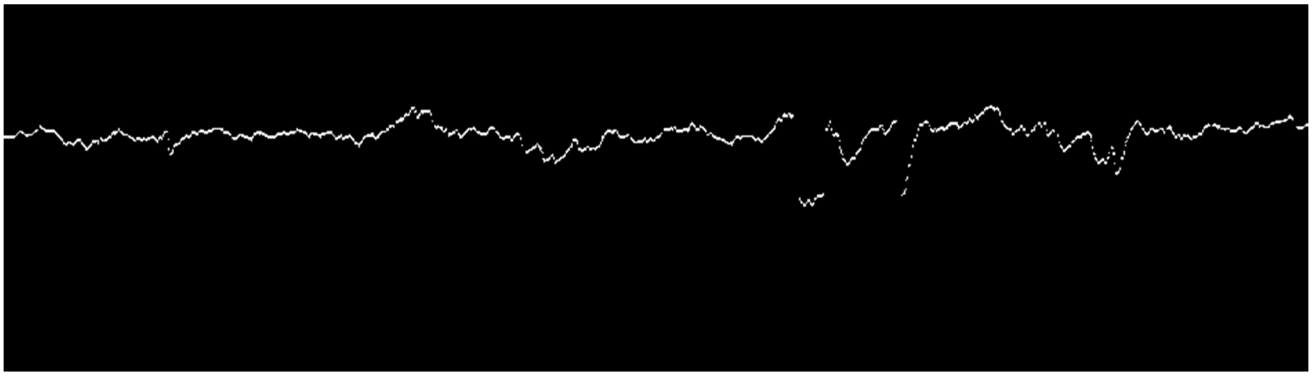


FIGURE 6. The thresholded image by the Otsu method.

Otsu thresholding is a method for determining the threshold value automatically based on the intensity histogram of the image. It assumes that the image has two classes of pixels: foreground and background. The basic idea was to find the threshold that maximizes the between-class variance, which is a measure of the separability of the two classes.

The algorithm works as follows:

1) Compute the histogram of the input image.

2) Normalize the histogram to obtain a probability density function.

3) Compute the cumulative sum and the cumulative mean of the normalized histogram.

4) Compute the global mean of the image.

5) Compute the between-class variance for all possible threshold values, using the formula:

$$\sigma^2(t) = \frac{(\mu_T.P_T - \mu_1)^2}{(P_T.(1 - P_T))} \quad (1)$$

where μ_T is the global mean, μ_1 is the mean of the foreground class up to the threshold t, P_T is the probability of the foreground class up to the threshold t, and $(1 - P_T)$ is the probability of the background class up to the threshold t. The optimal threshold value is the one that maximizes the between-class variance. Once the optimal threshold value is found, it can be used to segment the input image into foreground and background regions. Pixels with intensity values above the threshold are considered foreground pixels, while pixels with intensity values below the threshold are considered background pixels. After thresholding of Fig. 5, Fig. 6 is obtained. As seen in Fig. 5, the signal is shown with the white pixels in the dark background.

D. FINAL SIGNAL CONSTRUCTION

For obtaining the final FHR signal, for each column of the image as shown in Fig. 6, the row numbers of the white pixels are determined and the median of the row numbers was set as the signal value for the considered column. The obtained final FHR signal and its superimposed illustration in Fig. 4 are given in Fig. 7. As seen in Fig. 7, some parts of the signal are missing and a sliding window of size 300 sample length is used to complete the missing part of a signal by employing the median operation. Besides, a spline interpolation approach is used further tune the completed data accordingly. In Fig. 8 (a) and (b), the completion of the missing data points with moving median and the spline interpolation approaches are given.

As observed in Fig. 8 (a), while the moving median approach completes the missing data coarsely, the spline interpolation approach produces more tuned samples as shown in Fig. 8 (b).

After completion of the missing signal samples, all signals were resampled to 5120 sample length for making the length of the signals constant. Thus, the final signal of $y(t)$ is the amplitude of the signal in beats per minute (BPM) scale.

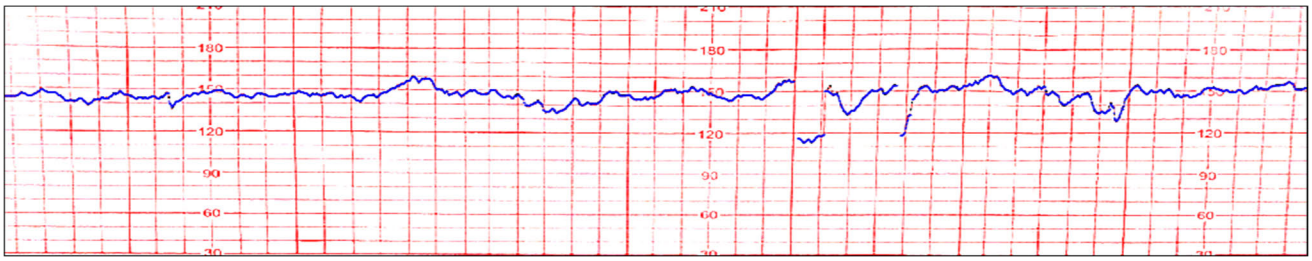
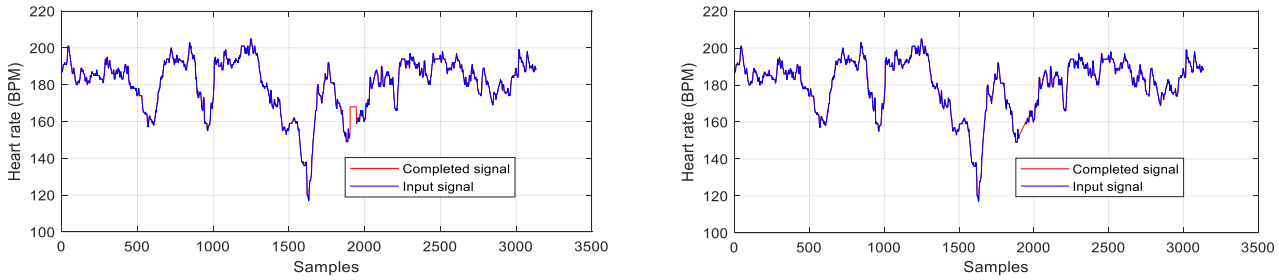


FIGURE 7. The obtained final FHR signal.



(a) Moving median (b) Spline interpolation

FIGURE 8. The missing data completion (a) Moving median approach, (b) Spline interpolation approach.

E. EMPIRICAL MODE DECOMPOSITION

Empirical Mode Decomposition (EMD) is a signal processing technique that decomposes a non-stationary signal into a set of Intrinsic Mode Functions (IMFs). The method was developed by Huang et al. [19] in 1998 and has since been widely used in various fields such as signal processing, finance, and neuroscience. The EMD method works by iteratively decomposing a signal into a set of. An IMF is extracted by identifying all the local extrema (maxima and minima) in the signal and interpolating between them using cubic splines. The interpolated extrema form the upper and lower envelopes of the IMF, which are then subtracted from the signal to obtain the first IMF. The process was then repeated on the residual signal until a stopping criterion is met. Let's consider a signal $x(t)$, the EMD algorithm works as follows:

1) Identify all the local maxima and minima in the signal $x(t)$. Let these be denoted as $maxima = \{m_1, m_2, \ldots, m_N\}$ and $minima = \{n_1, n_2, \ldots, n_N\}$, respectively.
2) Interpolate between the maxima and minima using cubic splines to obtain the upper and lower envelopes, respectively:
   - Upper envelope: $u(t) = spline(maxima)$
   - Lower envelope: $l(t) = spline(minima)$
3) Compute the mean of the envelopes:
   - Mean envelope: $m(t) = (u(t) + l(t))/2$
4) Subtract the mean envelope from the original signal to obtain the first IMF:
$$IMF1(t) = x(t) - m(t) \tag{2}$$
5) Repeat steps 1-4 on the residue signal ($IMF1(t)$) until a stopping criterion is met (e.g., a preset number of IMFs are extracted, or the residue becomes a monotonic function).
6) The final decomposition of the signal $x(t)$ is given by:
$$x(t) = IMF1(t) + IMF2(t) + \ldots + IMFk(t) + R(t) \tag{3}$$
where k is the total number of IMFs extracted and $R(t)$ is the residue signal. The EMD algorithm satisfies two criteria for each IMF:
   1. The IMF is a mono-component signal with a well-defined instantaneous frequency.
   2. The difference between the number of extrema and zero crossings is at most one.

The EMD algorithm is iterative, and the number of IMFs extracted depends on the complexity of the signal. The resulting IMFs can be used for further analysis, such as trend analysis or noise reduction. However, it should be noted that the EMD algorithm has some limitations, such as the possibility of mode mixing and the sensitivity to noise.

F. INSTANTANEOUS FREQUENCY AND SPECTRAL ENTROPY

Instantaneous frequency (IF) is a measure of the frequency of a signal at a specific point in time [20]. It can be calculated from the time-varying phase of the signal as follows:

- Let $x(t)$ be a signal with complex analytic representation $z(t) = x(t) + j * y(t)$, where j is the imaginary unit.
- The analytic signal can be written in polar form as $z(t) = A(t) * e^{(j\varphi(t))}$, where $A(t)$ is the instantaneous amplitude and $\varphi(t)$ is the instantaneous phase of the signal.
- The instantaneous frequency of the signal is given by the time derivative of the phase: $IF(t) = (\partial\varphi(t))/\partial t$

Spectral entropy (SE) is a measure of the complexity of a signal's frequency spectrum [21]. Spectral entropy is a measure of the distribution of power in the frequency domain and does not provide any information about the temporal dynamics of the signal. To capture the temporal dynamics, one can compute the spectral entropy over short time intervals using a sliding window approach. In this case, the spectral entropy was computed for each window, resulting in a time-varying measure of the complexity of the frequency spectrum. It is related to the entropy of the power spectral density (PSD) of the signal and is given by:

- Let $X(f)$ be the Fourier transform of a signal $x(t)$ with power spectral density $P(f) = |X(f)|^2$.
- The spectral entropy $H(f)$ is defined as: $H(f) = -1 \int P(f) * log2\,(P(f))\,df$

where the integral is taken over the frequency range of interest. The spectral entropy ranges from 0 to $log2(N)$, where N is the number of frequency bins in the PSD. A value of 0 indicates a completely flat spectrum, while a value of $log2(N)$ indicates a maximally complex spectrum with equal energy distributed across all frequency bins.

G. RELIEFF FEATURE SELECTION

The definition of the feature selection process is the removal of unnecessary, redundant, and noisy features from the original data set and the selection of the most significant features or optimum subset [22]. The Relief algorithm [22], a feature weighting technique, has been enhanced with ReliefF [23], [24]. Noisy and incomplete data cannot be handled by the original Relief method. Moreover, this approach can only be used to resolve binary classification issues. These issues can be dealt with and multi-class problems can be resolved using the ReliefF technique. While calculating the weights of the features, the ReliefF method substitutes the Manhattan distance for the Euclidean distance used in the original Relief algorithm. The Manhattan distance is used by the ReliefF method to first determine the k nearest hits H_j and k nearest misses $M_j(C)$ for the randomly chosen instance R_i. The weight vector $W[A_i]$ is then updated depending on the values of R_i, hitting H_j and missing $M_j(C)$ for all features A.

$$W\,[A_i] = W\,[A_i] - D_{A_i,R_i,H_j} + P_{A_i,R_i,H_j} \tag{4}$$

where D_{A_i,R_i,H_j} and P_{A_i,R_i,H_j} are defined as follows;

$$D_{A_i,R_i,H_j} = \sum_{j=1}^{k} diff\,\left(A_i, R_i, H_j\right)/m.k \tag{5}$$

$$P_{A_i,R_i,H_j} = \sum_{C\neq Class(R_i)} \frac{\left[\frac{P(C)}{1-P(class(R_i))} \sum_{j=1}^{k} diff\,\left(A_i, R_i, M_j\,(C)\right)\right]}{m.k} \tag{6}$$

where $W\,[A_i]$ represents is the weight value for the i_{th} feature $A.P\,(C)$ is the prior probability of class C, m is the process cycle, and $diff$ defines the Manhattan distance.

H. DATA DESCRIPTION

The dataset was constructed by the authors and all ethical issues were approved. The CTG dataset consists of 78 measurements (39 normal and 39 abnormal), each consisting of a series of 20-minute FHR and UC recordings, as well as the classification of the fetal state as either normal or abnormal. The measurements were obtained from 78 pregnant women who were admitted to the Department of Obstetrics and Gynecology at the Erzurum Regional Training and Research Hospital in Erzurum, Turkiye between December 2022 and January 2023. The recordings were obtained using a device called the Sonicaid System 8002, which was connected to the pregnant woman's abdomen using two transducers. Fig. 9 shows sample FHR signals for normal and abnormal classes, respectively. While the up row shows the normal FHR signals, the down row depicts the abnormal FHR signals.

IV. EXPERIMENTAL ANALYSIS

Experiments were conducted on a computer having an Intel i7 microprocessor and 64 GB Ram. All coding was carried out on MATLAB software. The experimental works were conducted based on 10-fold cross-validation criteria and average accuracy, sensitivity, specificity, and F1-score metrics were used for performance evaluation metrics [25]. The confusion matrix is a table used to evaluate the performance of a classification model. It consists of four elements: true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). TP is the model that correctly predicted a positive instance. FP is the model that predicted a positive instance when it was negative. TN is the model that correctly predicted a negative instance. And, FN is the model that predicted a negative instance when it was positive. Thus, accuracy is formulated as (TP + TN) / (TP + TN + FP + FN) the proportion of all correct predictions. Sensitivity is defined as TP / (TP + FN) the proportion of actual positive instances that were correctly identified by the model. Specificity is defined as TN / (TN + FP) the proportion of actual negative instances that were correctly identified by the model. Lastly, F1-score is defined as 2 * (precision * recall) / (precision + recall) - the harmonic mean of precision and recall, where precision = TP / (TP + FP) and recall = TP / (TP + FN). As mentioned earlier SVM classifier was used. For finding the optimal parameters of the SVM classifier, a hyperparameter tuning procedure was adopted for the SVM classifier. As the number of samples in normal and abnormal classes was low, each signal was divided into 512 samples length for data augmentation. Thus, a total of 780 signals were obtained for the classification part of the proposed study. A five-level EMD was employed and in Fig. 10, the decomposition levels were given for a normal and abnormal class. Columns a and b show the normal and abnormal classes respectively. The first row shows the input FHR signals and the other rows show the decomposition levels 1 to 5, respectively. For each decomposition level, the IF and SE features were extracted. Thus for each FHR signal, a 1290-length feature vector was obtained.

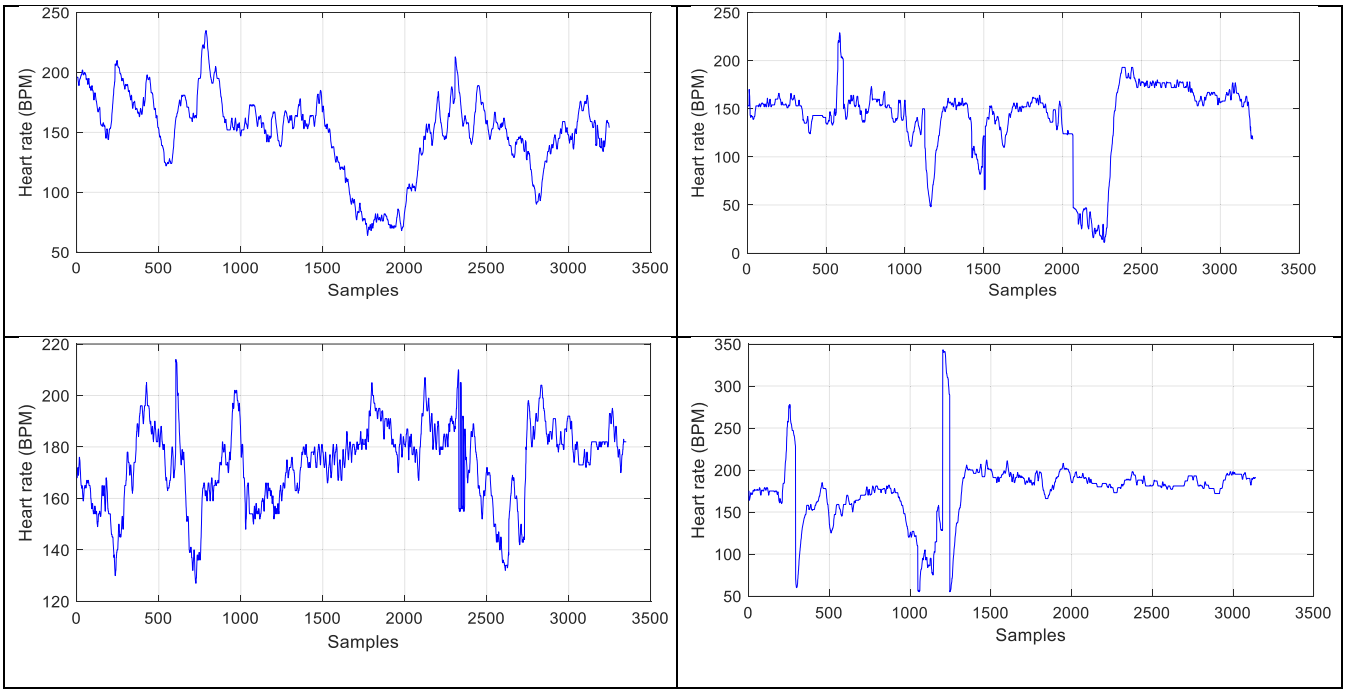


FIGURE 9. The sample FHR signals for normal and abnormal classes.

TABLE 1. The performance evaluation metrics for the proposed method.

	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1-Score (%)
SVM	99.62	99.74	99.48	99.62

TABLE 2. The performance evaluation metrics for the other classifiers.

	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1-Score (%)
DT	97.94	97.69	98.21	97.94
KNN	98.71	98.71	98.71	98.71
BC	98.33	98.46	98.20	98.33
NN	99.1	99.23	98.97	99.10

For ReliefF-based feature selection, the number of nearest neighbors value was set to 50 and the selected number of features was set to 500 [26]. These values were determined during the experimental works. In Fig. 11, the deviation of the average accuracy against the number of selected features was given. As seen in Fig. 11, the highest accuracy score was obtained around the number of features 500. Fig. 12 shows the minimum classification error plot of the optimized SVM classifier.

TABLE 3. The performance evaluation metrics for the CTU-UHB dataset.

	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1-Score (%)
Proposed method	99.70	96.42	100	98.18

The training was completed after 30 iterations. The best hyperparameters were obtained in the first iteration. The expected minimum classification errors were around zero except for the 4th iteration. The selected features were normalized to zero mean and unit variance procedure. The SVM training was run with a linear kernel and the training time was 206.48 seconds. The box constraint level was 0.0331. The hyperparameters were optimized with Bayesian optimization.

The obtained cumulative confusion matrix was given in Fig. 11. As seen in Fig. 13, only 3 samples were misclassified and a total of 777 samples were correctly classified. The Average accuracy was 99.62%.

The mentioned performance evaluation metrics were given in Table 1. As seen in Table 1, the sensitivity, specificity, and the F1-Score values were 99.74 %, 99.48%, and 99.62%, respectively.

Other classifiers namely decision tree (DT), k-nearest neighbor (KNN), Bayesian classifier (BC), and Neural networks (NN) also applied to the same task and the obtained results were given in Table 2 [27], [28], [29].Various classifiers were used to show the capability of the extracted and selected features in discrimination of the normal and abnormal fetus based on the CTG signals.

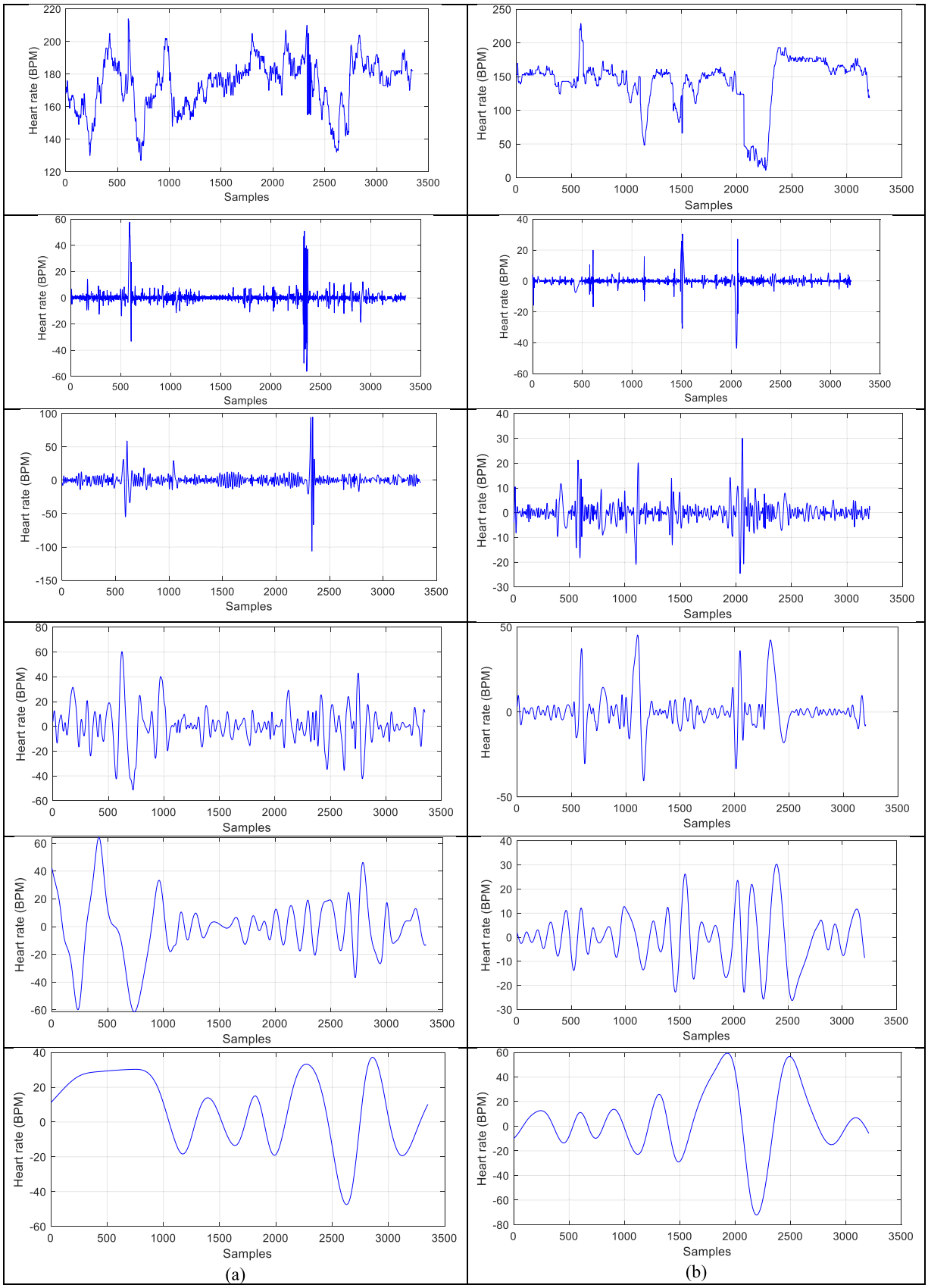


FIGURE 10. 5 level EMD for normal and abnormal classes, (a) Normal class and (b) Abnormal class.

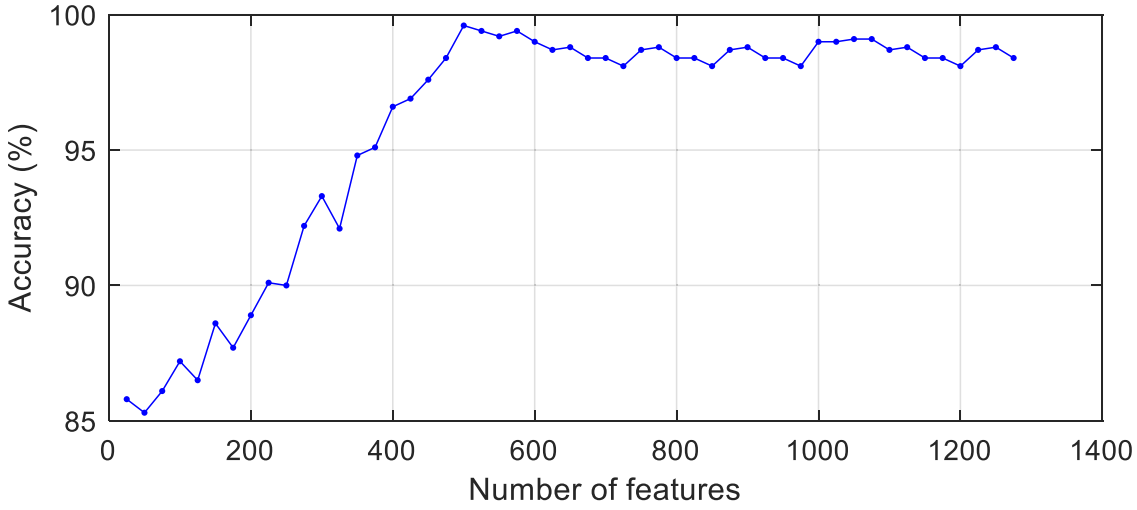


FIGURE 11. The accuracy against a number of selected features.

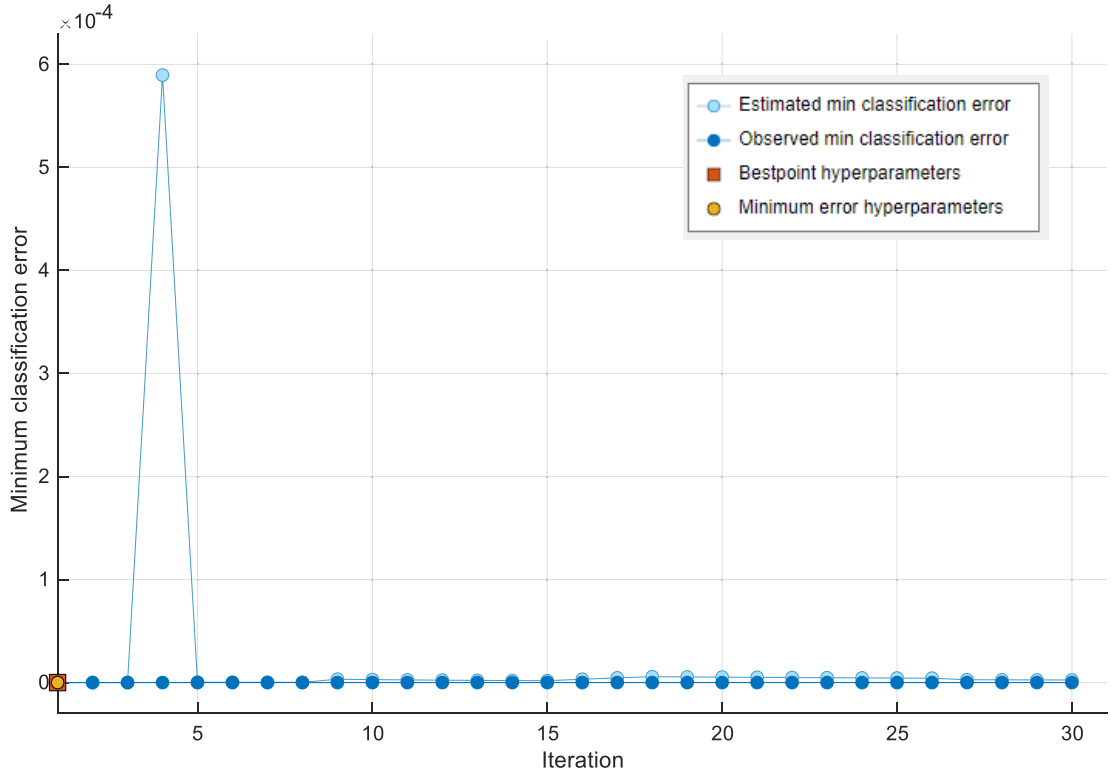


FIGURE 12. The minimum classification error plot for the optimized SVM model.

The hyper-parameters of these classifiers were also determined with Bayesian optimizations. As seen in Table 2, the accuracy scores were in the range of 97.94% and 99.1%. The NN classifier produced the highest accuracy score among the other determined classifiers. Similarly, the sensitivity scores were in the range of 97.69% and 99.23%. The specificity and F1-score values were also in the range of 98.21% and 98.97%, and 97.94% and 99.10%, respectively. From Table 2, it can be inferred that the NN classifier yielded better classification performance than the DT, KNN, and BC classifiers. Besides, it is worth mentioning that the SVM classifier outperformed the determined other classifiers.

Besides, the well-known CTU-UHB [11] dataset was also used for performance validation of the proposed method. The dataset contains 552 of the CTG samples and the length of the signal samples is 7200. The experiments were conducted based on the 10-fold-cross-validation test and obtained average results were given in Table 3. As seen in Table 3, a 99.70%

TABLE 4. Summary of comparison with state-of-art models.

Authors	Year	Dataset	Classifier	Validation Method	Accuracy
Ge et al. [9]	2019	CTU-UHB [11]	Deep learning	Hold out (75:25)	92.15 %
Rana et al. [10]	2020	CTU-UHB [11]	Ensemble of ML	Hold out (80: 20)	98.40 %
Chudáček et al. [11]	2014	CTU-UHB [11]	SVM	Hold out (75:25)	87.30%
Chudáček et al. [12]	2017	CTU-UHB [11]	ML	Hold out (75:25)	88.80%
Ajirak et al. [30]	2022	CTU-UHB [11]	Ensemble of classifiers	Leave two out cross-validation	83.70 %
Bursa et al. [31]	2017	CTU-UHB [11]	CNN	Hold out (80: 20)	94.10 %
Comert et al. [32]	2019	CTU-UHB [11]	CNN	Hold out (75:25)	93.32 %
Zhao et al. [33]	2019	CTU-UHB [11]	CNN	Hold out (80:20)	98.34 %
Daydulo et al. [34]	2022	CTU-UHB [11]	CNN	Hold out (80: 20)	98.70 %
Parvathavarthine et al. [35]	2020	CTU-UHB [11]	CNN	Cross-validation (5-fold)	94.63 %
Frasch et al. [36]	2021	Own dataset	CNN	Hold out (80:20)	93.60 %
Proposed method	2023	Own dataset	SVM	Cross-validation(10 fold)	99.62 %
Proposed method	2023	CTU-UHB [11]	SVM	Cross-validation(10 fold)	99.70%

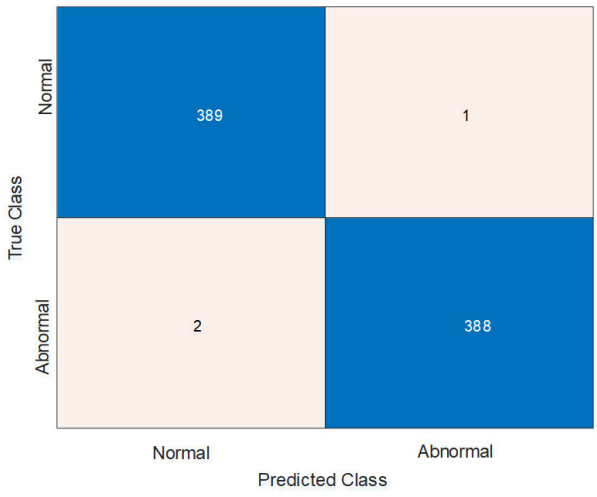


FIGURE 13. Cumulative confusion matrix for SVM classifier.

average accuracy score, 96.42% sensitivity, 100% specificity and 98.18% F1 score were produced by the proposed method.

V. RESULTS AND DISCUSSIONS

In this paper, a novel approach was presented for printing CTG paper digitization and abnormality detection based on digitized FHR signals. For printing CTG paper digitization, an image processing-based approach was used and for abnormality detection, machine learning was considered. The proposed method was quite efficient for both printing CTG paper digitization and the detection of abnormalities. A novel printing CTG dataset was used in this study and satisfied performance was obtained.

Table 4 shows a comparison of the proposed method with various studies that have been conducted so far. Ge et al. [9] used deep learning for CTG signal classification and obtained a 92.5% accuracy score. Rana et al. [10] used an ensemble of ML classifiers for normal and abnormal CTG signal classification and obtained a 98.40% accuracy score. In the works of Chudáček et al. [11], [12], authors used ML approaches for the classification of the CTG signals and obtained 87.3% and 88.8% accuracy scores, respectively. Ajirak et al. [30] used an ensemble of various classifiers with a publically available dataset for the detection of the abnormalities in CTG signals Authors used to leave two of our cross-validation approaches

in their experiments and obtained an 83.7% accuracy score.

Bursa et al. [31], Comert et al. [32], and Zhao et al. [33] used quite similar approaches for CTG signal classifications. All authors used the Continuous wavelet transform and short-time Fourier transform to convert the CTG signal into a CTG image and pre-trained convolutional neural networks (CNN) models were fine-tuned for classification purposes. These authors used the CTU-UHB dataset and obtained 94.1%, 93.32%, and 98.34%, respectively. Duydulo et al. [34] also used a CNN approach for CTG signal classification and obtained a 98.7% accuracy score. Parvathavarthine et al. [35] and Frasch et al. [36] also used CNN approach in their experiments with CTG signal classification and obtained 94.63% and 93.60 % accuracy scores respectively. Finally, in the last row of Table 4, the achievement of the proposed method on CTU-UHB dataset was given. As seen, a 99.7% average accuracy score was obtained with the proposed method which was the highest one that has been produced so far.

VI. CONCLUSION

This paper demonstrates the effectiveness of using image processing and machine learning techniques for the automatic digitization and classification of CTG recordings as normal or abnormal. A novel approach is developed where image processing algorithms were used for signal digitization and a set of features was extracted from the CTG recordings to achieve high levels of sensitivity and specificity in the abnormality diagnosis of fetal distress. The results of the study suggest that computerized systems for fetal distress diagnosis have the potential to improve the accuracy and consistency of fetal monitoring, which could lead to better outcomes for both the mother and the baby. Furthermore, the use of machine learning algorithms for fetal distress diagnosis can reduce the workload of obstetricians, allowing them to focus on more cases that are complex.

By the developed system, the CTG data may be diagnosed, increasing the likelihood of early intervention for the pregnant woman and potentially lowering the risk of difficulties for both mother and fetus. The diagnosis of this application will also help medical practitioners save time. It can assist midwives and their doctors in monitoring expectant patients while they are doing their internship training (clinical practices). It will eliminate any potential delays in the interpretation of the CTG for young midwives and doctors who have just entered the field. Hence, it aims to lower rates of sickness and death in both mothers and infants.

The main limitation of the proposed study is the form of the input CTG printing images. A well-designed scanning procedure should be carried out to standardize the CTG printing images. Besides, the manual cropping procedure can be seen as another limitation of the proposed method.

REFERENCES

[1] A. Pasarica, D. Nemescu, H. Costin, and C. Rotariu, "Automatic analysis of cardiotocographic recordings for fetal acidosis study," *Med.-Surgical J.*, vol. 121, no. 1, pp. 206–214, 2017.

[2] A. Alonso-Betanzos, M. Cabrero-Canosa, and V. Moret-Bonillo, "Automated analog-to-digital conversion of graphical cardiotocographic records," *J. Clin. Eng.*, vol. 20, no. 1, pp. 57–65, Jan. 1995.

[3] T. Frusca, M. Soregaroli, A. Valcamonico, L. Scalvi, R. Bonera, and U. Bianchi, "Effect of betamethasone on computerized cardiotocographic parameters in preterm growth-restricted fetuses with and without cerebral vasodilation," *Gynecol. Obstetric Invest.*, vol. 52, no. 3, pp. 194–197, Feb. 2001, doi: 10.1159/000052972.

[4] M. F. Helguera, D. P. Misra, W. Janjindamai, R. Suarez, and M. A. Arriaga, "Digital analysis of fetal heart rate tracings: A systematic review," *J. Perinatal Med.*, vol. 44, no. 1, pp. 49–58, 2016.

[5] B. O. Verburg, C. J. M. de Groot, and R. H. Stigter, "Computer analysis of intrapartum fetal heart rate patterns compared to visual cardiotocography: A retrospective analysis," *Acta Obstetricia Gynecologica Scandinavica*, vol. 97, no. 4, pp. 457–465, 2018, doi: 10.1111/aogs.13281.

[6] C. S. da Costa, G. Greisen, and T. Austin, "Is near-infrared spectroscopy clinically useful in the preterm infant?" *Arch. Disease Childhood Fetal Neonatal Ed.*, vol. 100, no. 6, pp. F558–F561, 2015, doi: 10.1136/archdischild-2014-307919.

[7] A. Gandjour, "Cost-effectiveness of sofosbuvir in hepatitis C genotype 1 infection in Germany: A reanalysis of published results," *PLoS ONE*, vol. 15, no. 10, Oct. 2020, Art. no. e0236543, doi: 10.1371/journal.pone.0236543.

[8] D. Ayres-de-Campos et al., "SisPorto 2.0: A program for automated analysis of cardiotocograms," *J. Maternal-Fetal Neonatal. Med.*, vol. 29, no. 2, pp. 219–226, 2016, doi: 10.1002/1520-6661(200009/10)9:5<311::AID-MFM12>3.0.CO;2-9.

[9] D. Wu, Y. Zhang, D. Zhu, S. Wang, and M. Shen, "A channel calibration algorithm based on isolated scatterers for multi-channel HRWS-SAR," *IEEE Access*, vol. 7, pp. 135665–135677, 2019, doi: 10.1109/ACCESS.2019.2941203.

[10] R. Zeng, Y. Lu, S. Long, C. Wang, and J. Bai, "Cardiotocography signal abnormality classification using time-frequency features and ensemble cost-sensitive SVM classifier," *Comput. Biol. Med.*, 2021, doi: 10.1016/j.compbiomed.2021.104218.

[11] V. Chudáèek, J. Andén, S. Mallat, P. Abry, and M. Doret, "Scattering transform for intrapartum fetal heart rate characterization and acidosis detection," in *Proc. 35th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Osaka, Japan, 2013, pp. 2898–2901, doi: 10.1109/EMBC.2013.6610146.

[12] V. Chudáèek et al., "Low dimensional manifold embedding for scattering coefficients of intrapartum fetale heart rate variability," in *Proc. 36th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Chicago, IL, USA, 2014, pp. 6373–6376, doi: 10.1109/EMBC.2014.6945086.

[13] A. Heinonen, M. Gissler, J. Paavonen, A.-M. Tapper, and M. Jakobsson, "Risk of preterm birth in women with cervical intraepithelial neoplasia grade one: A population-based cohort study," *Acta Obstetricia Gynecologica Scandinavica*, vol. 97, no. 2, pp. 135–141, Feb. 2018, doi: 10.1111/aogs.13256.

[14] Z. Cömert, A. Şengür, Y. Akbulut, Ü. Budak, A. F. Kocamaz, and S. Güngör, "A simple and effective approach for digitization of the CTG signals from CTG traces," *IRBM*, vol. 40, no. 5, pp. 286–296, Oct. 2019.

[15] Z. Cömert, A. Şengür, Y. Akbulut, Ü. Budak, A. F. Kocamaz, and V. Bajaj, "Efficient approach for digitization of the cardiotocography signals," *Phys. A, Stat. Mech. Appl.*, vol. 537, Jan. 2020, Art. no. 122725.

[16] A. Sbrollini, A. Agostinelli, I. Marcantoni, M. Morettini, L. Burattini, F. Di Nardo, S. Fioretti, and L. Burattini, "ECTG: An automatic procedure to extract digital cardiotocographic signals from digital images," *Comput. Methods Programs Biomed.*, vol. 156, pp. 133–139, Mar. 2018.

[17] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. SMC-9, no. 1, pp. 62–66, Feb. 1979.

[18] Y. Olmez, A. Sengur, G. O. Koca, and R. V. Rao, "An adaptive multilevel thresholding method with chaotically-enhanced Rao algorithm," *Multimedia Tools Appl.*, vol. 2022, pp. 1–27, Jan. 2022.

[19] N. E. Huang, K. Hu, A. C. C. Yang, H.-C. Chang, D. Jia, W.-K. Liang, J. R. Yeh, C.-L. Kao, C.-H. Juan, C. K. Peng, J. H. Meijer, Y.-H. Wang, S. R. Long, and Z. Wu, "On Holo–Hilbert spectral analysis: A full informational spectral representation for nonlinear and non-stationary data," *Phil. Trans. Roy. Soc. A, Math., Phys. Eng. Sci.*, vol. 374, no. 2065, Apr. 2016, Art. no. 20150206, doi: 10.1098/rsta.2015.0206.

[20] U. Budak, V. Bajaj, Y. Akbulut, O. Atila, and A. Sengur, "An effective hybrid model for EEG-based drowsiness detection," *IEEE Sensors J.*, vol. 19, no. 17, pp. 7624–7631, Sep. 2019.

[21] H. Uyanık, S. T. A. Ozcelik, Z. B. Duranay, A. Sengur, and U. R. Acharya, "Use of differential entropy for automated emotion recognition in a virtual reality environment with EEG signals," *Diagnostics*, vol. 12, no. 10, p. 2508, Oct. 2022.
[22] K. Kira and L. R. Aaai. (1992). *The Feature Selection Problem: Traditional Methods and a New Algorithm*. Accessed: Jun. 9, 2022. [Online]. Available: https://www.aaai.org/Library/AAAI/1992/aaai92-020.php
[23] I. Kononenko, "Estimating attributes: Analysis and extensions of RELIEF," in *Proc. ECML*, 1994, pp. 171–182.
[24] K. Demir, A. R. I. Berna, and F. Demir, "Detection of brain tumor with a pre-trained deep learning model based on feature selection using MR images," *FIRAT Univ. J. Experim. Comput. Eng.*, vol. 2, no. 1, pp. 23–31, 2023.
[25] D. Sengur, "EEG, EMG and ECG based determination of psychosocial risk levels in teachers based on wavelet extreme learning machine autoencoders," *Politeknik Dergisi*, vol. 25, no. 3, pp. 985–989, 2021.
[26] D. Şengür, "Investigation of the relationships of the students' academic level and gender with COVID-19 based anxiety and protective behaviors: A data mining approach," *Turkish J. Sci. Technol.*, vol. 15, no. 2, pp. 93–99, 2020.
[27] D. Sengür and M. Turhan, "Prediction of the action identification levels of teachers based on organizational commitment and job satisfaction by using k-nearest neighbors method," *Turkish J. Sci. Technol.*, vol. 13, no. 2, pp. 61–68, 2018.
[28] A. Ari, "Multipath feature fusion for hyperspectral image classification based on hybrid 3D/2D CNN and squeeze-excitation network," *Earth Sci. Inform.*, vol. 2023, pp. 1–17, Jan. 2023.
[29] F. Demir, A. Sengur, A. Ari, K. Siddique, and M. Alswaitti, "Feature mapping and deep long short term memory network-based efficient approach for Parkinson's disease diagnosis," *IEEE Access*, vol. 9, pp. 149456–149464, 2021.
[30] M. Ajirak, C. Heiselman, J. G. Quirk, and P. M. Djuric, "Boost ensemble learning for classification of CTG SIGNALS," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2022, pp. 1316–1320.
[31] M. Bursa, A. Holzinger, M. Renda, and S. Khuri, Eds., *Information Technology in Bio- and Medical Informatics* (Lecture Notes in Computer Science), vol 10443. Cham, Switzerland: Springer, 2017, doi: 10.1007/978-3-319-64265-9_9.
[32] Z. Cömert and A. F. Kocamaz, "Fetal hypoxia detection based on deep convolutional neural network with transfer learning approach," in *Proc. Comput. Sci. Conf.*, vol. 763, 2019, pp. 239–248, doi: 10.1007/978-3-319-91186-1_25.
[33] Z. Zhao, Y. Deng, Y. Zhang, Y. Zhang, X. Zhang, and L. Shao, "DeepFHR: Intelligent prediction of fetal acidemia using fetal heart rate signals based on convolutional neural network," *BMC Med. Informat. Decis. Making*, vol. 19, no. 1, pp. 1–15, Dec. 2019.
[34] Y. D. Daydulo, B. L. Thamineni, H. K. Dasari, and G. T. Aboye, "Deep learning based fetal distress detection from time frequency representation of cardiotocogram signal using Morse wavelet: Research study," *BMC Med. Informat. Decis. Making*, vol. 22, no. 1, p. 329, Dec. 2022.
[35] K. Parvathavarthine and R. Balasubramanian, "Optimized residual convolutional learning neural network for intrapartum maternal-embryo risk assessment," *Eur. J. Mol. Clin. Med.* vol. 7, no. 11, pp. 2985–3006, 2020.
[36] M. G. Frasch, S. B. Strong, D. Nilosek, J. Leaverton, and B. S. Schifrin, "Detection of preventable fetal distress during labor from scanned cardiotocogram tracings using deep learning," *Frontiers Pediatrics*, vol. 9, pp. 1–8, Dec. 2021, doi: 10.3389/fped.2021.736834.

SIBEL ÖZTÜRK received the B.Sc. degree in midwifery from Ankara University, Turkey, in 2002, and the M.Sc. degree in public health nursing and the Ph.D. degree in public health nursing from Atatürk University, Turkey, in 2012 and 2022, respectively. She became a Research Assistant with the Health Science Faculty, Atatürk University, in 2004, where she is currently an Assistant Professor with the Health Science Faculty. Her research interests include woman health, lactation period, and birth.

SAFIYE AĞAPINAR ŞAHIN received the B.Sc., M.Sc., and Ph.D. degrees in midwifery from Sivas Cumhuriyet University, Turkey, in 2001, 2011, and 2021, respectively. She became a Lecturer with the Health Science Faculty, Atatürk University, in 2004, where she is currently an Assistant Professor with the Health Science Faculty. Her research interests include women health, birth, and pregnancy.

AYŞE NUR AKSOY received the Assoc. Prof. Dr. and Prof. Dr. degrees in obstetrics and gynecology from the University of Health Sciences, Erzurum Regional Training and Research Hospital, Erzurum, Turkey, in 2015 and 2022, respectively. She is currently a Doctoral Professor with the Department of Obstetrics and Gynecology, University of Health Sciences. Her research interests include general obstetrics, obstetric Doppler, general gynecology, and infertility.

BERNA ARI received the B.Sc. degree in computer technology, and the M.Sc. and Ph.D. degrees in electrical and electronics engineering from Firat University, Turkey, in 2013, 2017, and 2022, respectively. Her research interests include data augmentation, machine learning, and artificial intelligence.

ALEX AKINBI received the Ph.D. degree in computer security from Edge Hill University, U.K., in 2016. He is currently a Senior Lecturer in computer forensics and cyber security with Liverpool John Moores University, U.K. He has served for several years in academic positions and worked with leading industry players in cyber security. He has a long history of working in different areas of computer forensics and security as a security researcher, an ethical hacker, a security consultant, and a professional trainer. His research interests include mobile, the IoT, cloud forensics and security, machine learning and artificial intelligence application, cyber security and vulnerability research, reverse engineering, and exploit development, where he has contributed several journal articles, conference papers, and book chapters.