

RESEARCH

Open Access



# Optimizing poultry audio signal classification with deep learning and burn layer fusion

Esraa Hassan<sup>1</sup>, Samar Elbedwehy<sup>2</sup>, Mahmoud Y. Shams<sup>1</sup>, Tarek Abd El-Hafeez<sup>3,4\*</sup> and Nora El-Rashidy<sup>1</sup>

\*Correspondence:  
tarek@mu.edu.eg

<sup>1</sup> Department of Machine Learning and Information Retrieval, Faculty of Artificial Intelligence, Kafrelsheikh University, Kafr El-Sheikh 33516, Egypt

<sup>2</sup> Department of Data Science, Faculty of Artificial Intelligence, Kafrelsheikh University, Kafrelsheikh 33516, Egypt

<sup>3</sup> Department of Computer Science, Faculty of Science, Minia University, El-Minia 61519, Egypt

<sup>4</sup> Computer Science Unit, Deraya University, El-Minia, Egypt

## Abstract

This study introduces a novel deep learning-based approach for classifying poultry audio signals, incorporating a custom Burn Layer to enhance model robustness. The methodology integrates digital audio signal processing, convolutional neural networks (CNNs), and the innovative Burn Layer, which injects controlled random noise during training to reinforce the model's resilience to input signal variations. The proposed architecture is streamlined, with convolutional blocks, densely connected layers, dropout, and an additional Burn Layer to fortify robustness. The model demonstrates efficiency by reducing trainable parameters to 191,235, compared to traditional architectures with over 1.7 million parameters. The proposed model utilizes a Burn Layer with burn intensity as a parameter and an Adamax optimizer to optimize and address the overfitting problem. Thorough evaluation using six standard classification metrics showcases the model's superior performance, achieving exceptional sensitivity (96.77%), specificity (100.00%), precision (100.00%), negative predictive value (NPV) (95.00%), accuracy (98.55%), F1 score (98.36%), and Matthew's correlation coefficient (MCC) (95.88%). This research contributes valuable insights into the fields of audio signal processing, animal health monitoring, and robust deep-learning classification systems. The proposed model presents a systematic approach for developing and evaluating a deep learning-based poultry audio classification system. It processes raw audio data and labels to generate digital representations, utilizes a Burn Layer for training variability, and constructs a CNN model with convolutional blocks, pooling, and dense layers. The model is optimized using the Adamax algorithm and trained with data augmentation and early-stopping techniques. Rigorous assessment on a test dataset using standard metrics demonstrates the model's robustness and efficiency, with the potential to significantly advance animal health monitoring and disease detection through audio signal analysis.

**Keywords:** Poultry audio classification, Deep learning, Convolutional neural networks, Digital audio signal processing, Disease detection, Sensitivity, Specificity

## Introduction

Recent advances in technology enable the exploitation of animals' acoustic communication for automatic health monitoring, offering real-time, objective, and cost-effective alternatives to manual inspection methods [1–3].

Despite the increasing interest in exploiting animal vocalizations for automated health monitoring purposes, few openly available datasets support the development of such systems [4]. Specifically, for poultry production, only limited resources documenting healthy versus diseased states' vocal patterns are public [5–7]. This study utilizes a labeled audio dataset of healthy and unhealthy chicken vocalizations from various African poultry breeds. The dataset contains 346 WAV recordings categorized as healthy, noisy, or unhealthy. Each category contains 139, 86, and 121 instances, respectively, lasting between 5 and 60 s. Selected sound segments in the unhealthy folder include chicken cough, snoring, and rale sounds, representing symptoms of respiratory distress. At the same time, the Noise folder contains background noises and poultry bird activities such as feeding and pecking one another.

Analyzing animal vocalizations involves extracting relevant features from raw audio data. Various representations capture different aspects of audio signals, e.g., Mel-frequency cepstral coefficients (MFCC), chromograms, and autocorrelations. Subsequent processing typically entails dimensionality reduction and normalization techniques to feed downstream machine learning algorithms [8].

Deep neural networks constitute popular choices for pattern recognition problems requiring end-to-end mapping between raw sensor data and desired outputs [9]. Several architectures achieve excellent performances on various image, video, audio, and text modalities [10, 11]. Among those, ResNet (short for residual network) gained popularity due to its skip connections mitigating vanishing gradient issues encountered during optimization [12]. Extensions of plain ResNets include bottleneck structures and dilated convolutions, boosting representational capacities while controlling computational costs [13].

However, despite their proven abilities, deep learning models remain sensitive to hyperparameter tuning and prone to memorizing random artifacts rather than genuine underlying patterns [14, 15]. Regularization schemes counteract overfitting tendencies, inducing prior distributions over model parameters and introducing stochasticity during optimization [16]. Examples include weight decay, dropouts, and adversarial perturbations added to input data. Recently, the latter idea inspired the introduction of so-called "Burn-in" layers, injecting structured noise directly inside neural networks to encourage more stable training and enhanced generalization properties [17].

Considering the scarcity of openly available datasets depicting healthy vs. unhealthy chicken vocalizations, this study proposes an original architecture harnessing the above-mentioned techniques to discriminate between the considered classes [18]. Starting from raw waveform representations, mel spectrograms serve as intermediate visual descriptors capturing spectral patterns. Then, two parallel pipelines process local vs. global features separately [19]. Local details pass through consecutive convolutional layers, while global trends proceed straight to the final merge point. Before merging, both paths apply separate temporal pooling stages to reduce dimensionality. Lastly, fully connected layers supported by dropouts produce probabilistic estimates reflecting the likelihood of observing either healthy or unhealthy cases [20].

This research compares the proposed architecture to baselines established on plain ResNets and simple combinations of convolutional and recurrent modules. Quantitative assessments rely on sensitivity, specificity, positive/negative predictive values, accuracy,

F1 scores, and Matthew's correlation coefficient. Visual inspections complement numerical analyses by scrutinizing learning curves and plotting error evolution across iterations. Obtained results shed light on the relative strengths and weaknesses of competing designs, guiding future improvements in automatic poultry health monitoring systems.

Overall, this work addresses the pressing need for intelligent and autonomous monitoring systems in agriculture, particularly poultry farming. By capitalizing on readily deployable sensors collecting vast amounts of multimedia data, the envisaged framework shall alert farmers about anomalous situations threatening their businesses' sustainability. Therefore, immediate actions tackling emerging threats become possible, limiting financial damages and safeguarding food security.

### **Problem statement**

Reliable and efficient monitoring of poultry health is crucial for the poultry industry. This study addresses the challenges in developing accurate and robust poultry sound classification models. Environmental factors can distort audio signals, making it difficult to extract relevant features for classification. Additionally, training stability issues in deep learning models can lead to unreliable predictions [21]. To enhance robustness, this study integrates a Burn Layer into ResNet-based architectures for improved training stability.

A dataset comprising 346 audio signal files is used, categorized as healthy, noisy, and unhealthy. Each file represents a distinct time frame, with carefully selected sound segments. The "noisy" category includes background noises from vehicles and human voices, as well as poultry bird activities. The "unhealthy" category contains sounds indicative of respiratory issues. All files are stored in .wav audio format [18].

Two architectures are implemented: a traditional architecture and a proposed work architecture. Both architectures feature convolutional and pooling layers, batch normalization, and activation functions. The proposed work architecture includes a Burn Layer, which adds Gaussian noise during training to improve robustness. It also incorporates a second input layer and a global average pooling layer.

The study evaluates the performance of these architectures to demonstrate the effectiveness of the Burn Layer in improving robustness and stability in poultry sound classification models. The aim is to establish a foundation for accurate and dependable poultry health monitoring systems, benefiting the poultry industry and animal welfare standards.

### **Research objective**

This paper aims to enhance the robustness of deep learning models for classifying poultry multimedia data by introducing novel modifications to existing architectures. The Burn Layer introduces controlled noise during training, significantly improving the model's stability and generalization by exposing it to noisy inputs. This innovative technique sets it apart from traditional regularization methods and offers a novel solution to the challenges of training robust deep learning models. The proposed architecture not only reduces the number of trainable parameters to 191,235 but also achieves outstanding classification metrics. This approach has significant potential applications in animal health monitoring and disease detection through audio signal analysis.

## Contributions

This paper proposes modifications to existing deep learning architectures for enhanced robustness in classifying poultry multimedia data. Specifically, it makes the following key contributions:

1. Introduction of a Burn Layer that randomly perturbs input data during training. This improves model stability and generalization by exposing it to noisy inputs.
2. Development of an end-to-end pipeline for audio-based poultry health status detection using deep learning. Previous work in this domain utilized traditional machine-learning techniques.
3. Comparison of performance between a standard CNN, ResNet with Burn Layer, and the proposed model integrating Burn Layers. The proposed model achieves superior results with 98.55% accuracy.
4. Detailed analysis of model optimization steps including data preprocessing, augmentation, training with early stopping, and learning rate scheduling.
5. Evaluation of models on important metrics like sensitivity, specificity, precision, etc. to provide a holistic view of classification capability.

This paper advances the field of poultry disease detection by developing an accurate deep learning-based approach and customized architecture incorporating techniques to enhance robustness. The openly available dataset also enables further research.

The remainder of this paper is structured as follows: “[Related work](#)” section provides a review of related work. “[Preliminaries](#)” section outlines the preliminaries which include the most common methodologies Burn Layer, ResNet, ResNext, DenseNet, and Wide ResNet. “[Proposed work](#)” section presents the proposed work; the results and analysis of the proposed ensemble model are presented in “[Result and experimental](#)” section. Finally, “[Discussions and limitations](#)” section concludes the paper, highlighting future research directions.

## Related work

### Previous approaches

Monitoring poultry traits is crucial for assessing environmental health conditions and making informed decisions [22, 23]. These aggregated pieces of information are utilized to ensure the welfare of poultry and make appropriate management choices [24]. Research focusing on changes in physiological traits can be employed to predict variations in vocalization patterns and detect various diseases [25, 26]. Poultry, being a homeothermic animal, generates and disperses heat to maintain a constant body temperature [18, 27]. Several methods have been developed to monitor the health status of poultry, one of which involves monitoring their body temperature [18, 29]. Fluctuations in body temperature can indicate stress or pathological conditions. Therefore, temperature monitoring plays a significant role in determining the health status of poultry. Infrared thermoregulation (IRT) is a technique commonly used to measure poultry temperature [28, 29]. Additionally, visual observation is employed to identify sick chickens [27, 30–34]. Recent studies suggest that vocalization and

abnormal sounds emitted by sick chickens can serve as significant indicators of their health status [35, 36]. In one study, Qunitee et al. [37] developed a hybrid model for chicken monitoring using a decision tree (DT). The system was based on visual input from 15 chickens monitored over 72 h, achieving a classification accuracy of 84.8%. When audio inputs were considered, the classification accuracy improved to 86.1%. Another study [38] developed a detection system aiming to classify chick calls based on Deep Learning models. The study explored three different chick breeds, analyzing zero-crossing rate and short-time crossing rate to identify the endpoints of chick calls in the audio signals. The results showed that the ResNet model achieved the highest accuracy of 83%, while the gated recurrent network (GRU) achieved an accuracy of 90%.

### Challenges

Despite these advancements, several challenges persist in effectively monitoring poultry health through audio signals. The complexity of accurately detecting and classifying various health indicators such as stress, diseases, and other pathological conditions through sound remains high. Noise in the environment often interferes with the accuracy of the models, making it difficult to distinguish between normal and abnormal sounds. Additionally, the diversity in vocalization patterns among different breeds and individual chickens adds another layer of complexity. This necessitates the development of sophisticated models that can generalize well across different conditions and environments.

### Proposed solutions

To address these challenges, various innovative approaches have been proposed. For instance, a study [39] utilized a ResNet model to classify Newcastle disease among poultry, which affects both health and production. The study utilized audio signals from 35 chickens and implemented multi-window spectral filtering and high-filtering techniques to reduce the impact of noise. The processed model achieved an average accuracy of 91.06% for infected and healthy chicken classes. In another study [40], an audio-based system was developed to detect chicken stress during their first weeks of life. The system monitored the birds' sounds, identified stress, and improved any conditions that may have arisen. The study concluded that pre-recorded audio signals could be used with different classifiers and at different frame levels. Using four classifiers at a 1000 ms frame level, accuracies ranged from 63 to 83%. Additionally, authors in [41] proposed an audio-based system to detect various types of vocalization in chickens, including chirping, peeping, and begging. The system analyzed the audio signals, applied feature engineering, and utilized joint-time–frequency scattering (JTFS) for feature extraction to accurately identify the different types of vocalizations. Another innovative solution [42] involved developing a system capable of detecting chicken sneezing in noisy environments. The researchers built a model based on audio data for sneeze and non-sneeze classification, aggregating 763 audio segments from 51 chickens. The system achieved a performance of 88.4% and 66.7% in terms of sensitivity and precision, respectively. Furthermore, a study focused on the early detection of influenza in chickens by analyzing audio signals and extracting sound features using Mel-frequency cepstral coefficients (MFCC) to classify healthy and infected chickens [27]. The model achieved accuracy

ranging from 84 to 90%. These studies collectively demonstrate the potential of audio signals to provide significant indications about the health status of poultry. Table 1 provides a summary of the current related work.

## Preliminaries

### The burn layer

The Burn Layer acts as a custom layer in a neural network that introduces a form of noise during training [44]. Initialization involves setting parameters, and controlling the intensity of the noise introduced by the layer, to a default value of 0.2 [45]. During the forward pass, when the layer is called with inputs, it checks if the model is in training mode [46–48]. By adding a special layer that enforces a selective burning process during training. The idea is to make the model more robust by training it on slightly perturbed data, akin to dynamic data augmentation. The burn intensity parameter allows control over the strength of this effect.

To represent the Burn Layer mathematically, given an input tensor  $X$  of shape  $(N, T, C)$ , the Burn Layer operation can be expressed as in Eq. (1)

$$\text{Burn Layer}(X) = X + \text{burn\_intensity} \times Z \quad (1)$$

where  $Z$  is a tensor of the same shape as  $X$ , containing random noise sampled from a normal distribution  $N(0, I)$  and scaled by the burn intensity. This operation adds random noise to the input tensor  $X$ , scaled by the burn intensity. The outputs include a digital audio signal represented as a sequence of samples. Steps involve calculating the total number of samples, initializing the digital audio signal array, sampling the continuous audio signal, quantizing the sampled values, normalizing the digital audio signal by converting quantized samples to a suitable digital representation (e.g., 16-bit integer), and scaling the values to fit within the dynamic range of the chosen representation. Finally, the digital audio signal array  $D$  is returned [45]. The mathematical formulation of the Burn Layer operates as follows:

- i) For a given layer  $L$  in the neural network, let  $h_L$  represent the output of the neurons in that layer. The Burn Layer modifies these outputs as in Eq. (2).

$$h'_L = h_L \odot m_L \quad (2)$$

where  $m_L$  is a mask vector of the same dimension as  $h_L$ , and  $\odot$  denotes the element-wise multiplication.

- ii) The mask  $m_L$  is updated based on a burning function  $f(h_L)$  that determines the likelihood of each neuron being burned as shown in Eq. (3).

$$m_L = m_L \odot \text{Burn}(h_L, \theta) \quad (3)$$

where  $\text{Burn}(h_L, \theta)$  is a function parameterized by  $\theta$  that progressively zeros out neurons as in Eq. (4).

$$\text{Burn}(h_L, \theta) = \begin{cases} 0 & \text{if } |h_L| < 0 \\ 1 & \text{otherwise} \end{cases} \quad (4)$$

**Table 1** Summarization of literature

Author	Dataset	Purpose	Algorithms	Results
[43]	22 h sound of unhealthy and healthy chicken via inoculating the SPF chickens and the H7N9 AV	Develop an early detection system for influenza in chickens by analyzing audio signals and extracting sound through the utilization of Mel-frequency cepstral coefficients (MFCC)	Support vector machine	Achieved an accuracy ranging from 84 to 90%
[32]	Images aggregate based on thermal camera	Develop a real-time system to monitor poultry by checking skin temperature by using a thermal camera	ANOVA test	The study concluded that Thermal imaging is a promising approach to quickly screen chickens for potential Influenza (HPAI) infection
[33]	450 image file	Propose a model to predict skin temperature based on the CNN model	Convolutional neural network	The combination of images from the thermal camera and CNN model provides accurate prediction
[34]	8067 image file	Developed CNN model that utilized to classify various chicken diseases	CNN	An average accuracy of 94.4
[37]	15 chickens for 72 h	Develop a hybrid model for disease monitoring and identification	Decision tree	Achieved performance up to 86.1 for classification accuracy
[38]	Aggregated audio signals within one day	Develop a detection system that utilizes Deep Learning models to classify chick calls	Deep learning models (GRU, ResNet-50)	ResNet model demonstrated accuracy at 83%, while the gated recurrent network (GRU) accuracy at 90%
[39]	Audio signals collected from 35 chickens	Build ResNet model categorizes Newcastle disease in poultry, based on multi-window spectral filtering and high filtering techniques	ResNet-50 model	average accuracy of 91.06% in distinguishing between infected and healthy chicken classes
[40]	Audio signals that were analyzed at 1000 MS frame level	Develop an audio-based system to identify chicken stress in the early weeks of their lives	Four deep learning classifiers	The average accuracies achieved ranged from 63 to 83%
[41]	Detecting and chick call candidates from full-length recordings	Propose an audio-based system aimed at identifying various types of vocalizations in chickens, such as chirping, peeping, and begging. This approach enabled the accurate differentiation of the different types of vocalizations	Using JTFS for precise feature extraction, then using deep learning for vocalization identification	Achieved an Average accuracy that ranged from 77 to 88%
[42]	Aggregating 763 audio segments from 51 chickens	Propose monitoring chicken sneezing in noisy environments. The system achieved a performance of	Deep learning models	The model achieved performance of 88.4% and 66.7% in terms of sensitivity and precision, respectively



where  $\theta$  is a threshold parameter that determines the sensitivity of the burning process.

iii) The overall loss function  $\mathcal{L}$  of the neural network might also incorporate a burning penalty to promote the burning of non-essential neurons as in Eq. (5).

$$\mathcal{L} = \mathcal{L}_{task} + \lambda \sum_L ||m| | \tag{5}$$

where  $\mathcal{L}_{task}$  is the original task-cross-entropy for classification), and  $\lambda$  is a regularization parameter controlling the strength of the burning penalty [49].

The comparison between the Burn Layer with traditional regularization techniques is investigated as follows:

i) *Dropout*: Dropout randomly sets a fraction of input units to zero during training, effectively reducing the network capacity and preventing overfitting. The Burn Layer, on the other hand, adds controlled noise to the input, maintaining the network’s full capacity but training it on slightly perturbed data as in Eq. (6).

$$h'_L = h_L \odot d_L \tag{6}$$

where  $d_L$  is a binary mask with each element being zero with probability  $p$ . Dropout aims to prevent co-adaptation of neurons by randomly omitting them during training. In contrast, the Burn Layer deterministically removes neurons based on their utility.

ii) *Weight Decay (L2 Regularization)*: This technique penalizes large weights by adding a regularization term to the loss function as shown in Eq. (7).

$$\mathcal{L} = \mathcal{L}_{task} + \lambda \sum_w w^2 \tag{7}$$

The Burn Layer does not directly influence the weights but rather the input data, ensuring the model learns to be robust to variations.

iii) *Data Augmentation*: Traditional data augmentation techniques create multiple modified copies of the training data. The Burn Layer dynamically perturbs the data during training, providing a similar effect but without the need for explicitly generating augmented datasets.

iv) *Batch Normalization*: Batch normalization standardizes the outputs of neurons to have zero mean and unit variance, followed by a learnable scaling and

v) shifting as shown in Eq. (8).

$$h'_L = \gamma \frac{h_L - \mu}{\sigma} h_L + \beta \tag{8}$$

where  $\mu$  and  $\sigma$  are the mean and standard deviation of the batch,  $\gamma$  and  $\beta$  are learnable parameters. Batch normalization primarily addresses internal covariate shift



and accelerates training, while the Burn Layer focuses on pruning non-contributive neurons [50].

**The ResNet model**

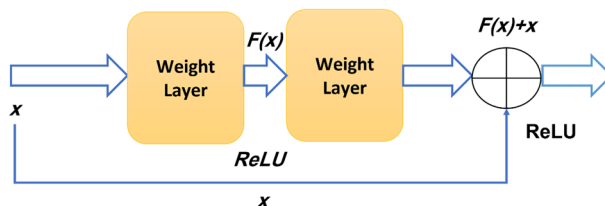
Residual Network (ResNet) is a deep learning model tailored for computer vision tasks, particularly designed to accommodate hundreds or thousands of convolutional layers [12]. Traditional CNN architectures struggled to scale to such depths due to the "vanishing gradient" issue, where gradients diminish with increasing layer depth, leading to suboptimal performance [51]. ResNet addresses this by introducing "skip connections," which allow the reuse of previous layer activations, thus mitigating the vanishing gradient problem [52].

By stacking multiple identity mappings and skipping layers during initial training, ResNet compresses the network into fewer layers, accelerating training [53]. Subsequently, during retraining, the skipped layers are expanded, enabling the network to explore more complex features of the input image. ResNet models typically skip two or three layers at a time, incorporating nonlinearity and batch normalization between them. Advanced versions, like HighwayNets, can dynamically determine skip weights to further optimize performance. Residual blocks form the core of the ResNet architecture, differing from older architectures like VGG16 [54], which relied on stacking convolutional layers with batch normalization and nonlinear activation layers [55]. While effective for a limited number of layers, subsequent research revealed the potential for improved performance with increased layer depth. ResNet’s simple yet effective approach of adding intermediate inputs to convolution blocks allows for deeper exploration of feature spaces, making it a powerful tool for computer vision tasks [56].

Figure 1 illustrates a typical residual block, which can be represented in Python code as output equal to  $F(x) + x$ , where  $x$  is the input to the block and the output from the previous layer, and  $F(x)$  represents the operations within the residual block. This technique, known as skip connections, facilitates smoother gradient flow during backpropagation, enabling networks to scale to significant depths, such as 50, 100, or even 150 layers, without suffering from the vanishing gradient problem. Importantly, skipping connections incurs no additional computational overhead [57]. This approach has gained widespread popularity and has been adopted in various neural network architectures beyond CNNs, including UNet and Recurrent Neural Networks (RNNs) [58].

**Optimization algorithms with CNN**

Sound recognition, a cornerstone in audio processing, encompasses tasks like audio classification and sound event detection [59]. Comprising various layers, each



**Fig. 1** The building block of the residual learning of ResNet architecture

assumes a distinct role in feature extraction and learning from input audio data [60]. Convolutional layers lie at the heart of the architecture, applying filters to detect patterns. Activation functions introduce non-linearities, while pooling layers reduce spatial dimensions, preserving crucial features. Fully connected layers process high-level features, and dropout mitigates overfitting by randomly deactivating neurons [61].

Optimization algorithms, such as Adam [62], Nadam [63], and Adamax [63], drive the training of sound recognition models. Adam, an amalgamation of AdaGrad [64] and RMSProp [65], adapts learning rates and maintains moving averages of gradients. Nadam integrates Nesterov's accelerated gradient, optimizing convergence [11]. Adamax, a streamlined variant of Adam, employs the max norm of gradients. These algorithms dynamically adjust learning rates, facilitating efficient parameter updates to minimize loss and enhance the performance of sound recognition systems.

### Dataset description

The dataset was collected from poultry birds that were purchased for 100 days. These birds were divided into groups at a research farm located at Bowen University. The birds have respiratory diseases that have many symptoms including cough, rale, and snoring. Some groups of birds received treatments while others did not. Afterward, the birds were separated and monitored in isolated environments [18]. To minimize noise interference, the microphones were placed away from the birds. Sound segments were recorded using 24-bit samples at a sampling rate of 96 kHz for 65 days.

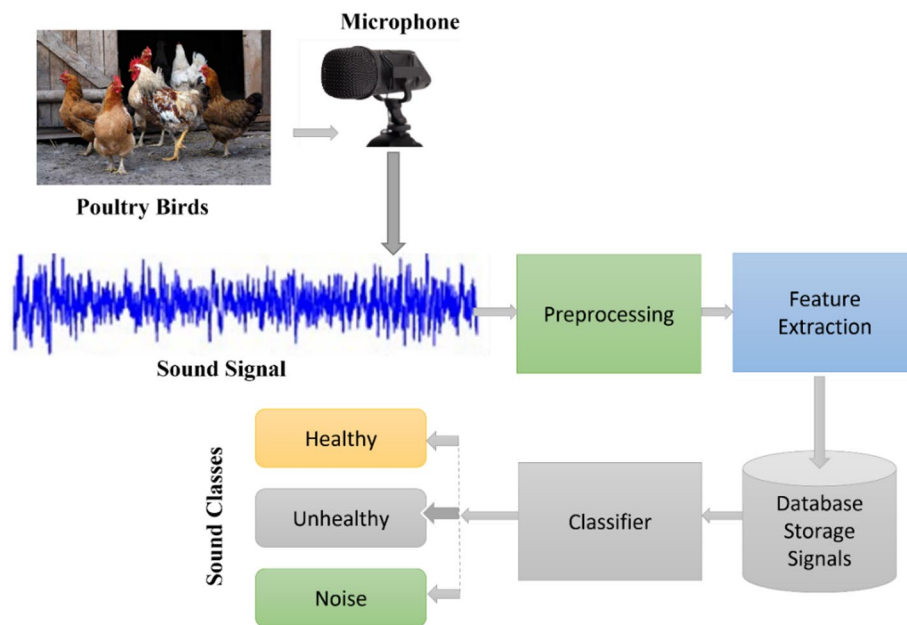
The dataset consists of 346 audio signals, which are categorized into three folders: "noisy," "healthy," and "unhealthy." The "healthy" folder contains 139 files, the "noisy" folder contains 86 audio files, and the "unhealthy" folder contains 121 audio files. The length of each audio signal ranges from 5 to 60 units. The selected noise segments include the sounds of moving vehicles, human voices, and other background noises. The selected segments in the "unhealthy" folder consist of cough, rale, and snore sounds. All files are stored in the .wav format [18] found in the following Mendeley link "<https://data.mendeley.com/datasets/zp4nf2dxbh/1>".

To analyze the audio signals, both frequency and time domain analyses were conducted. The power of the signals in the "healthy" and "unhealthy" categories was compared in the frequency domain, while the noise segments exhibited higher frequency content on the y-axis compared to the x-axis. The statistical analysis of the applied dataset including mean, standard deviation, skewness, kurtosis, median, range, interquartile range, and entropy are shown in Table 2.

To investigate further, the audio signal underwent analysis in both the time and frequency domains. Comparison was made between the power spectra of two signals, one representing healthy sounds and the other unhealthy sounds, within the frequency domain. In this comparison, it was observed that the normal sound exhibited a higher frequency content precisely around the y-axis. Conversely, the anomalous sound displayed a distinct spike of approximately 0.2 radians, indicating a noticeable deviation from the expected frequency distribution.

**Table 2** The statistical analysis of the applied poultry bird’s dataset

Statistic	Mean	Standard deviation	Skewness	Kurtosis	Median	Range	Interquartile range	Entropy
MFCC 1	- 287.98	12.52	0.76	0.47	- 290.41	62.63	6.84	4.50
MFCC 2	103.24	7.12	- 0.96	1.92	103.47	41.05	3.78	4.50
MFCC 3	24.85	6.97	- 0.73	0.29	26.36	31.48	5.64	4.45
MFCC 4	27.11	6.86	0.82	0.54	28.13	33.93	3.62	4.46
MFCC 5	13.24	5.52	0.16	0.53	13.22	24.26	4.03	4.41
MFCC 6	35.17	4.67	0.22	0.07	34.49	25.30	2.39	4.49
MFCC 7	4.17	4.65	0.40	0.27	3.70	21.50	3.93	- ∞
MFCC 8	2.95	4.98	- 0.19	0.26	2.71	28.00	3.20	- ∞
MFCC 9	15.54	5.09	0.19	- 0.77	15.77	21.41	4.64	4.45
MFCC 10	12.31	4.80	0.14	- 0.25	11.79	22.18	2.49	4.42
MFCC 11	- 10.25	4.58	0.69	0.23	- 10.71	21.86	2.64	- ∞
MFCC 12	13.16	4.15	0.22	0.36	12.92	22.65	2.66	4.45
MFCC 13	- 0.35	4.30	0.08	0.20	0.01	21.82	3.47	- ∞



**Fig. 2** The general structure of the proposed sound poultry recognition system

**Proposed work**

Figure 2 illustrates the overall architecture of the proposed model. The primary focus of this paper is on leveraging sound analysis algorithms for classifying chicken behavior within a farm environment. The sound signals undergo preprocessing, followed by feature extraction, and the extracted features are stored in a database [66]. Subsequently, these features are fed into a classifier, enabling the classification of sound signals into three categories: healthy, unhealthy, and noise. This approach aims to provide a systematic method for monitoring and assessing the well-being of chickens based on their vocalizations.

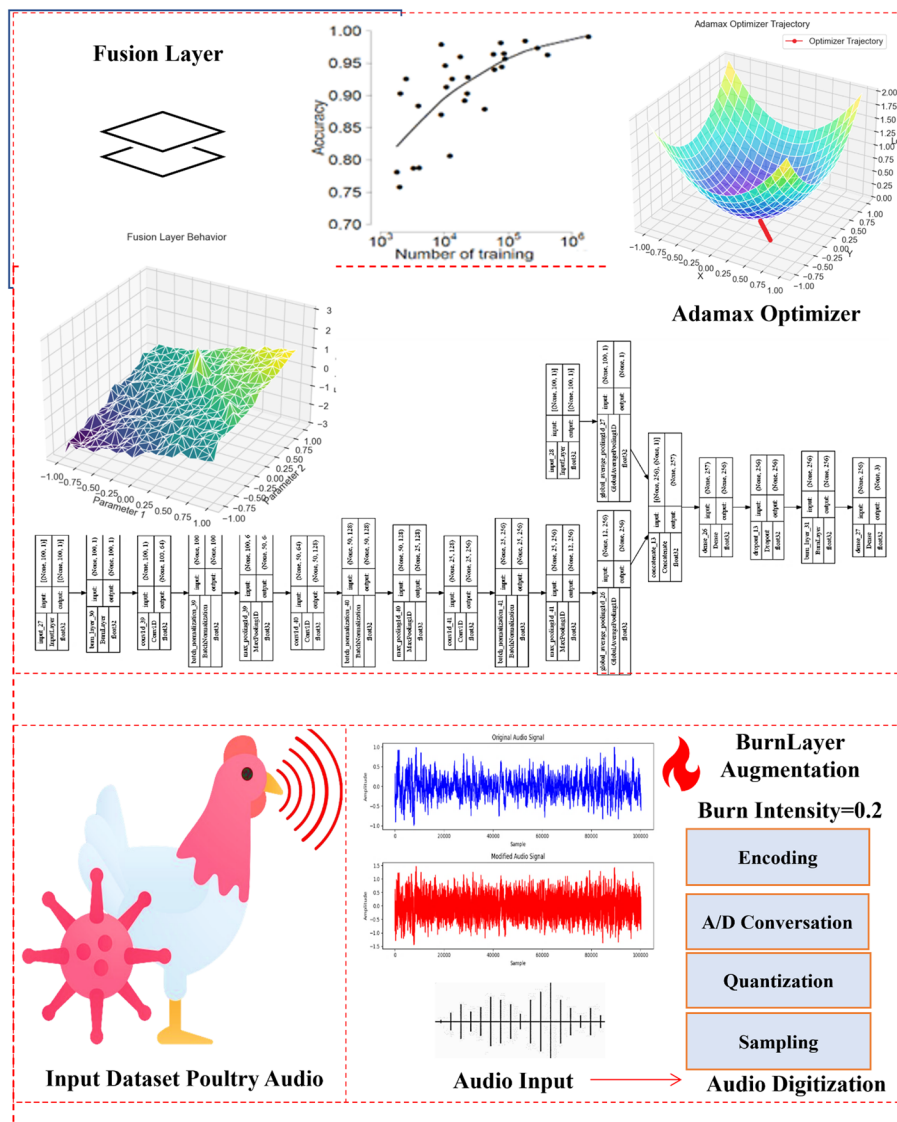
The proposed model presents the key steps to develop and evaluate a deep learning-based model for poultry audio classification that incorporates a Burn Layer for improved robustness. It takes in the raw training, validation, and test audio data as well as labels and pre-processes the data to create digital representations. It first defines a Burn Layer custom layer that randomly perturbs the input during training with a specified burn intensity parameter. This layer is applied to the input to expose it to noisy variations of the training samples. It then builds the classification model, starting with convolutional blocks comprising Conv1D, batch normalization, and max pooling layers. These layers extract powerful audio features from the input. A fusion layer is created by concatenating global average pooling outputs. Further, dense and dropout layers are added for classification. Crucially, another Burn Layer is integrated after the first dense layer for additional robustness. It is compiled using Adamax optimizer, categorical cross-entropy loss and trained over multiple epochs with data augmentation and early stopping. During training, validation loss is monitored to retain the best-performing weights. Once training is complete, it evaluates the trained model on the held-out test set to analyze performance metrics like accuracy, sensitivity, specificity, etc. This helps gauge how effectively the model with Burn Layers can classify poultry health from audio recordings.

Figures 3 and 4 outlines the proposed Burn Layer model. Our proposed Burn Layer model involves defining a custom layer responsible for adding controlled random noise to input data during training, thus improving robustness against fluctuations in input signals. To prepare the digital audio signal, we convert the continuous audio signal into a digital representation by determining the number of samples, initiating a zero-valued array, sampling the signal at equal intervals, transforming analog values into digital ones, and scaling them accordingly.

For designing the audio model architecture, we define input tensors, connect the Burn Layer to the input tensor with a specified burn intensity, stack sequential convolutional blocks—each containing a Conv1D layer, Batch Normalization layer, and MaxPooling1D layer—and configure activations and L2 regularization. Global average pooling compresses feature maps along the temporal dimension, allowing us to merge the aggregated features with another globally pooled input using a concatenation layer. We append fully connected layers—featuring a dense layer followed by a dropout layer and another Burn Layer with decreased burn intensity—culminating in the output layer activated with softmax for multi-class classification. Configuring the model appropriately includes setting the optimizer (Adamax), loss function (sparse categorical cross-entropy), and learning rate (0.001). For training, we implement early stopping, store the best model checkpoint based on validation loss, fix the learning rate, generate synthetic training data through data augmentation, divide input data into batches, and fine-tune model parameters using backpropagation. Once the model is trained, measuring its performance relies on computing the loss value and metric evaluations on held-back testing data. Our Burn Layer model provides enhanced robustness compared to conventional methods in diverse applications.

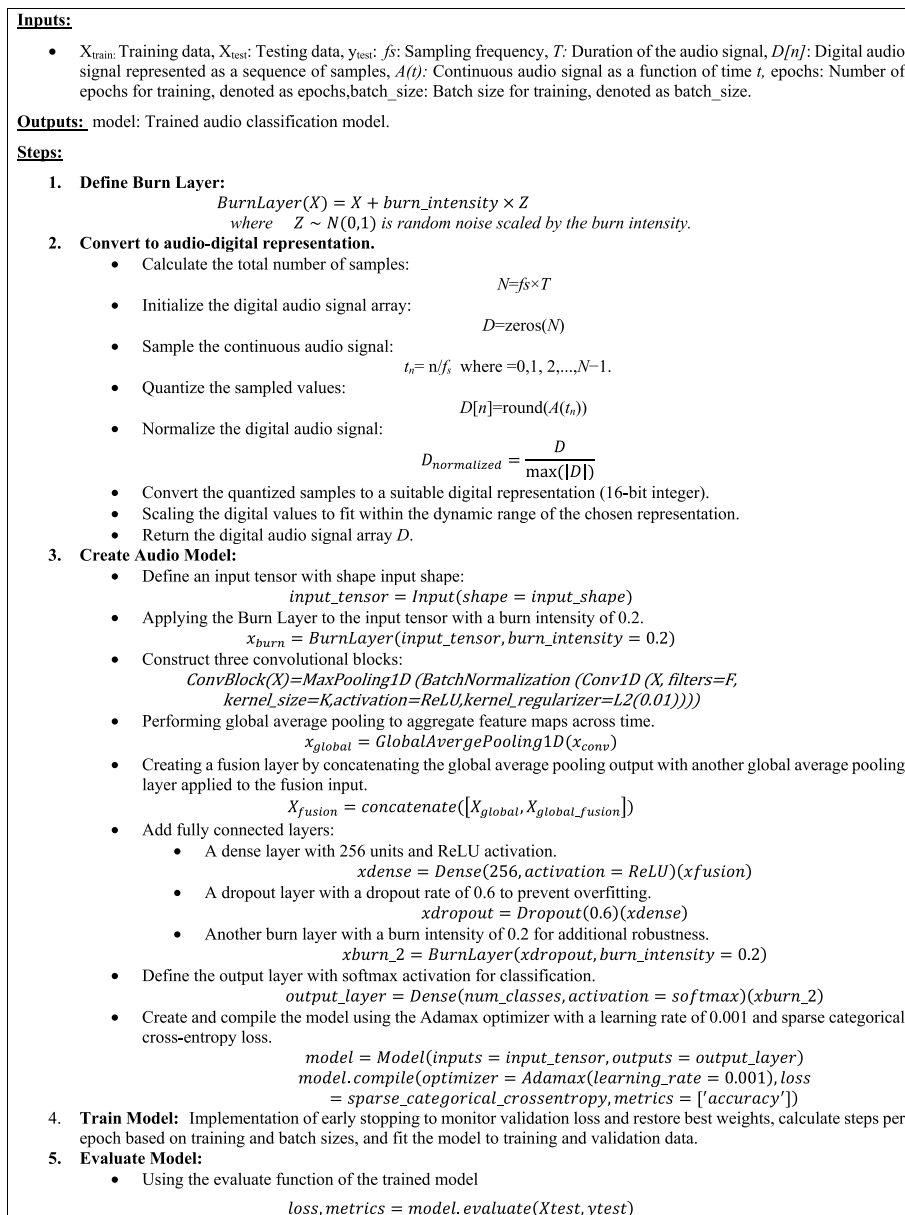
### **Data preprocessing**

The sound samples were framed and filtered as part of the sound signal preprocessing. Longer period nonstationary sound samples were then framed by a shifting



**Fig. 3** The Block diagram of the proposed model architecture

Hamming window to create a 10 to 30 s stationary signal. Table 3 details the architecture of the traditional convolutional neural network model used for comparison in the paper. The table lists each layer of the network, the output shape at each step, and the number of trainable parameters. The network takes an input of shape (None, 13, 1) representing the audio samples. It then applies a series of 1D convolutional layers interspersed with batch normalization, activation, and max pooling layers to extract features from the input audio. Three convolutional blocks are used, each containing a convolutional layer, batch normalization, and max pooling. This is followed by the global average pooling of feature maps across time. A fusion layer is created by concatenating outputs from two global average pooling layers. Finally, the network contains dense layers for classification. In total, the traditional architecture comprises 45



**Fig. 4** The Algorithm steps of the proposed model architecture

layers with a trainable parameter. The table provides a detailed overview of the network architecture to facilitate results comparison with the proposed model.

Table 4 outlines the architecture of the proposed model in this study. It improves upon the traditional architecture by incorporating a Burn Layer to enhance robustness. The input is first passed through the Burn Layer, which randomly perturbs the data during training. Then, similar to Table 3, it applies convolutional and max pooling layers in three blocks to extract features. A key difference is the introduction of a second input stream that is pooled separately and concatenated with the mainstream, forming a fusion layer. The network also contains dense layers for classification. Notably, another Burn Layer is added after the first dense layer. In total, the proposed model contains

**Table 3** The traditional architecture with a detailed description

Step	Layer name	Output shape	No. params	Step	Layer name	Output shape	No. params
1	Input	(None, 13, 1)	N/A	24	Convolutional layer 9	(None, 3, 256)	24
2	Conv1D	(None, 13, 64)	256	25	Batch normalization 7	(None, 3, 256)	25
3	Batch normalization	(None, 13, 64)	256	26	Batch normalization 8	(None, 3, 256)	26
4	Activation	(None, 13, 64)	0	27	Element-wise addition 3	(None, 3, 256)	27
5	Conv1D	(None, 13, 64)	12,352	28	Activation 5	(None, 3, 256)	28
6	Conv1D	(None, 13, 64)	128	29	Max pooling layer 3	(None, 1, 256)	29
7	Batch normalization	(None, 13, 64)	256	30	Convolutional layer 10	(None, 1, 512)	30
8	Batch normalization	(None, 13, 64)	256	31	Batch normalization 9	(None, 1, 512)	31
9	Element-wise addition	(None, 13, 64)	0	32	Activation 6	(None, 1, 512)	32
10	Activation	(None, 13, 64)	0	33	Convolutional layer 11	(None, 1, 512)	33
11	Max pooling1d	(None, 6, 64)	0	34	Convolutional layer 12	(None, 1, 512)	34
12	Conv1D	(None, 6, 128)	24,704	35	Batch normalization 10	(None, 1, 512)	35
13	Convolutional layer 5	(None, 6, 128)	49,280	36	Batch normalization 11	(None, 1, 512)	36
14	Convolutional Layer 6	(None, 6, 128)	8320	37	Element-wise addition 4	(None, 1, 512)	37
15	Batch normalization 4	(None, 6, 128)	512	38	Activation 7	(None, 1, 512)	38
16	Batch normalization 5	(None, 6, 128)	512	39	Second input layer	(None, 13, 1)	39
17	Element-wise addition 2	(None, 6, 128)	0	40	Global average pooling 1D layer 1	(None, 512)	40
18	Activation 3	(None, 6, 128)	0	41	Global average pooling 1D layer 2	(None, 1)	41
19	Max pooling layer 2	(None, 3, 128)	0	42	Global average pooling1d	(None, 512)	0
20	Convolutional layer 7	(None, 3, 256)	98,560	43	Global average pooling 1D	(None, 1)	0
21	Batch normalization 6	(None, 3, 256)	1024	44	Concatenate	(None, 513)	0
22	Activation 4	(None, 3, 256)	0	45	Dense	(None, 3)	1542
23	Convolutional layer 8	(None, 3, 256)	196,864		Total params: 1,748,806 Trainable params: 1,743,046 Non-trainable params: 5760		

19 layers with 191,235 trainable parameters—a more efficient architecture compared to Table 3. The Burn Layers aim to improve model stability during training by exposing it to variations in input. This table provides details of the modified architecture designed to classify poultry audio with improved robustness.

In Fig. 5, we observe the distribution of channels across the layers of a convolutional neural network (CNN). The number of channels in each layer is not uniform and varies



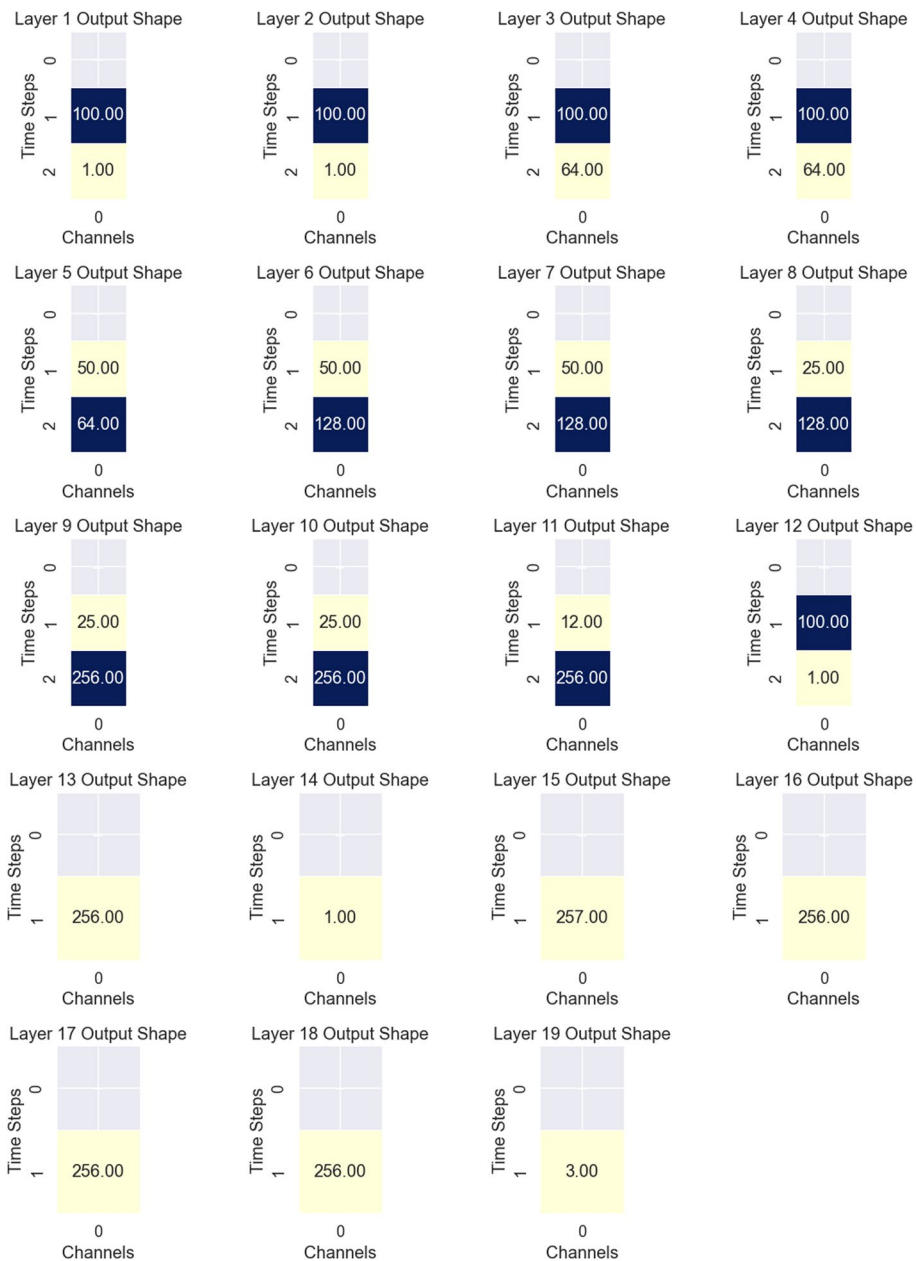
**Table 4** The proposed work architecture with a detailed description

Step	Layer name	Output shape	No. params
1	input_25 (InputLayer)	(None, 13, 1)	0
2	burn_layer (BurnLayer)	(None, 13, 1)	0
3	conv1d_144 (Conv1D)	(None, 13, 64)	256
4	batch_normalization_144	(None, 13, 64)	256
5	max_pooling1d_36	(None, 6, 64)	0
6	conv1d_145 (Conv1D)	(None, 6, 128)	24,704
7	batch_normalization_145	(None, 6, 128)	512
8	max_pooling1d_37	(None, 3, 128)	0
9	conv1d_146 (Conv1D)	(None, 3, 256)	98,560
10	batch_normalization_146	(None, 3, 256)	1024
11	max_pooling1d_38	(None, 1, 256)	0
12	input_26 (InputLayer)	(None, 13, 1)	0
13	global_average_pooling1d_24	(None, 256)	0
14	global_average_pooling1d_25	(None, 1)	0
15	concatenate_12 (Concatenate)	(None, 257)	0
16	dense_12 (Dense)	(None, 256)	66,048
17	dropout	(None, 256)	0
18	burn_layer_1 (BurnLayer)	(None, 256)	0
19	dense_13 (Dense)	(None, 3)	771
Total params	192,131	Trainable params	191,235
Non-trainable params		896	

depending on the layer's position relative to the output layer. Specifically, the initial layer contains 1 channel, followed by 64 channels in the second layer, and 128 channels in the third layer. The quantity of channels within a layer correlates with the number of filters applied within that layer. Filters play a pivotal role in feature extraction from input images. Consequently, layers with a higher number of channels possess an enhanced capacity to extract a more diverse range of features from the input image.

### Model evaluation

Following the completion of the training phase with an 80% training and 20% testing data split for our classification model, we meticulously evaluate its performance using established metrics. These measures encompass various facets of model performance in classification tasks. Accuracy gauges overall precision by comparing correctly predicted samples to the total, offering a broad measure of correctness. Precision focuses on the model's accuracy in identifying positive instances, calculated as the ratio of true positives to the total predicted positives. Recall (or sensitivity) evaluates the model's ability to capture positive instances by comparing true positives to the total actual positives. The F1 Score, a harmonic mean of precision and recall, provides a balanced assessment, crucial for scenarios with imbalanced class distributions. Additionally, loss, derived from an average cross-entropy error during training, indicates the model's ability to predict the correct class and guides optimization. AUC (Area Under the ROC Curve) measures the model's discriminative ability between classes, calculated based on the entire area under the curve plotting True Positive Rate vs. False Positive Rate. These metrics collectively offer a comprehensive view of a classification model's effectiveness, enabling assessment across various aspects of accuracy



**Fig. 5** The number of channels in each layer of a convolutional neural network (CNN)

and predictive capability. Common techniques, outlined in references and Eqs. (9)–(15), are utilized for the computation of sensitivity, specificity, precision, Negative Predictive Value (NPV), accuracy, F1-score, and Matthews Correlation Coefficient (MCC), respectively. [11, 67, 68].

$$\text{Recall} = \text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{9}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (10)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (11)$$

$$\text{Negative Predictive Value} = \frac{\text{TN}}{\text{TN} + \text{FN}} \quad (12)$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \quad (13)$$

$$\text{F1 - score} = 2 * \frac{(\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})} \quad (14)$$

$$\text{MCC} = \frac{\text{TP} \times \text{TN} - \text{FP} \times \text{FN}}{\sqrt{(\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN})}} \quad (15)$$

where TP, TN, FN, and FP are truly positive, true Negative, False Negative, and False Positive numbers respectively. n is the number of classes.

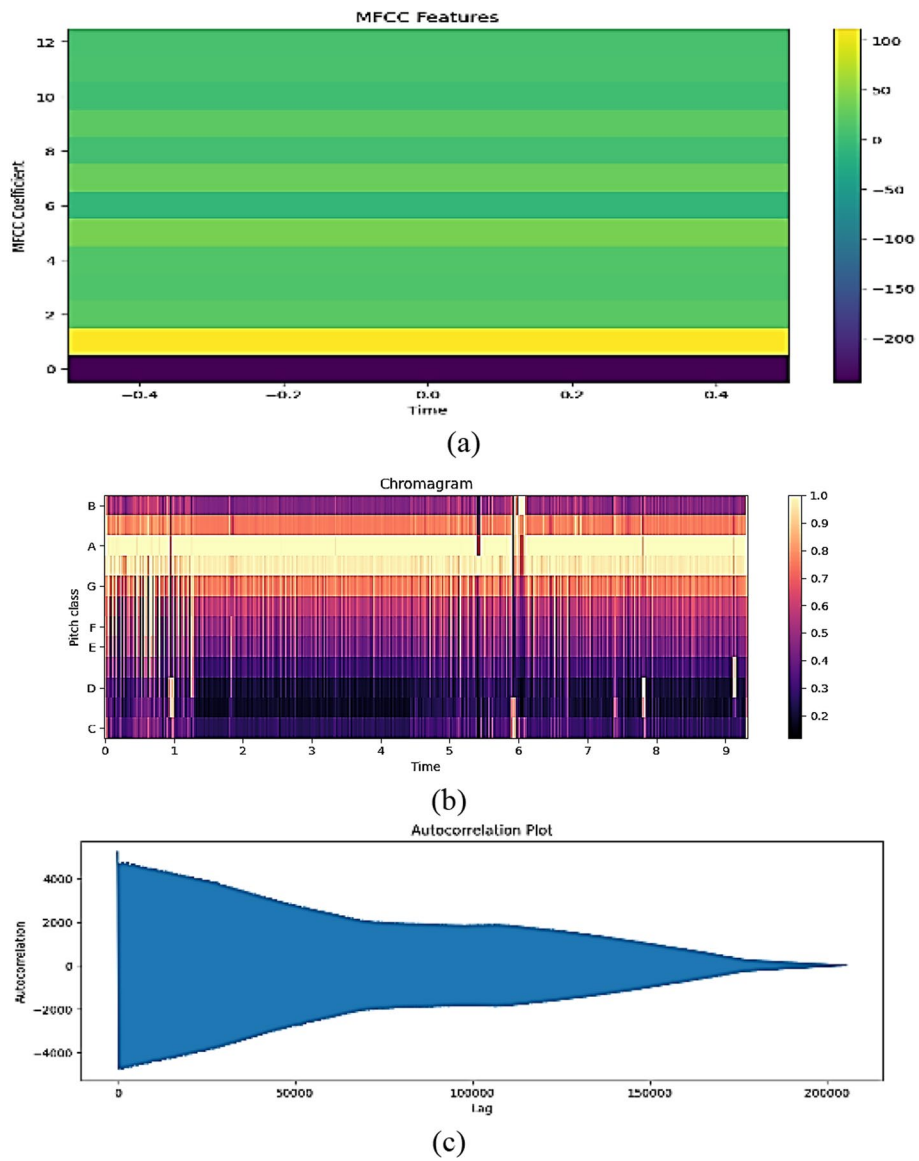
## Result and experimental

To evaluate the effectiveness of our machine learning framework, we conducted experiments in this section. The experiments were performed on a computer with a 3 GHz i5 processor, 8GB main memory, and a 64-bit Windows 10 operating system. We used the Python programming language to experiment.

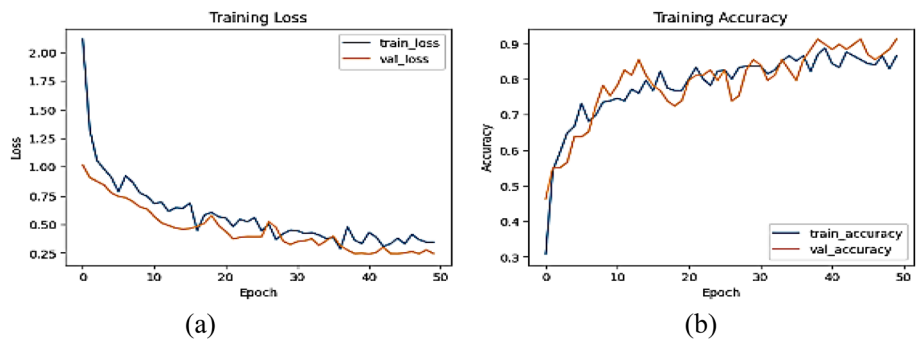
Firstly, we train our model using CNN and NADAM, and the resulting MFCC features are shown in Fig. 6a, while the chromogram that shows the relation between the time and pitch size is shown in Fig. 6b. The autocorrelation plot that shows the lags that vary from 0 to 200,000 is shown in Fig. 6c. The learning curve of the CNN + NADAM model is shown in Fig. 7.

Secondly, we utilized a model that undergoes training utilizing ResNet (50), Burn Layer, and NADAM, resulting in MFCC features depicted in Fig. 8a. Additionally, Fig. 8b illustrates the chromogram, elucidating the relationship between time and pitch size. Furthermore, Fig. 8c showcases the autocorrelation plot, exhibiting lags ranging from 0 to 100,000. The learning curve of the ResNet (50) + Burn Layer + NADAM model is presented in Fig. 9.

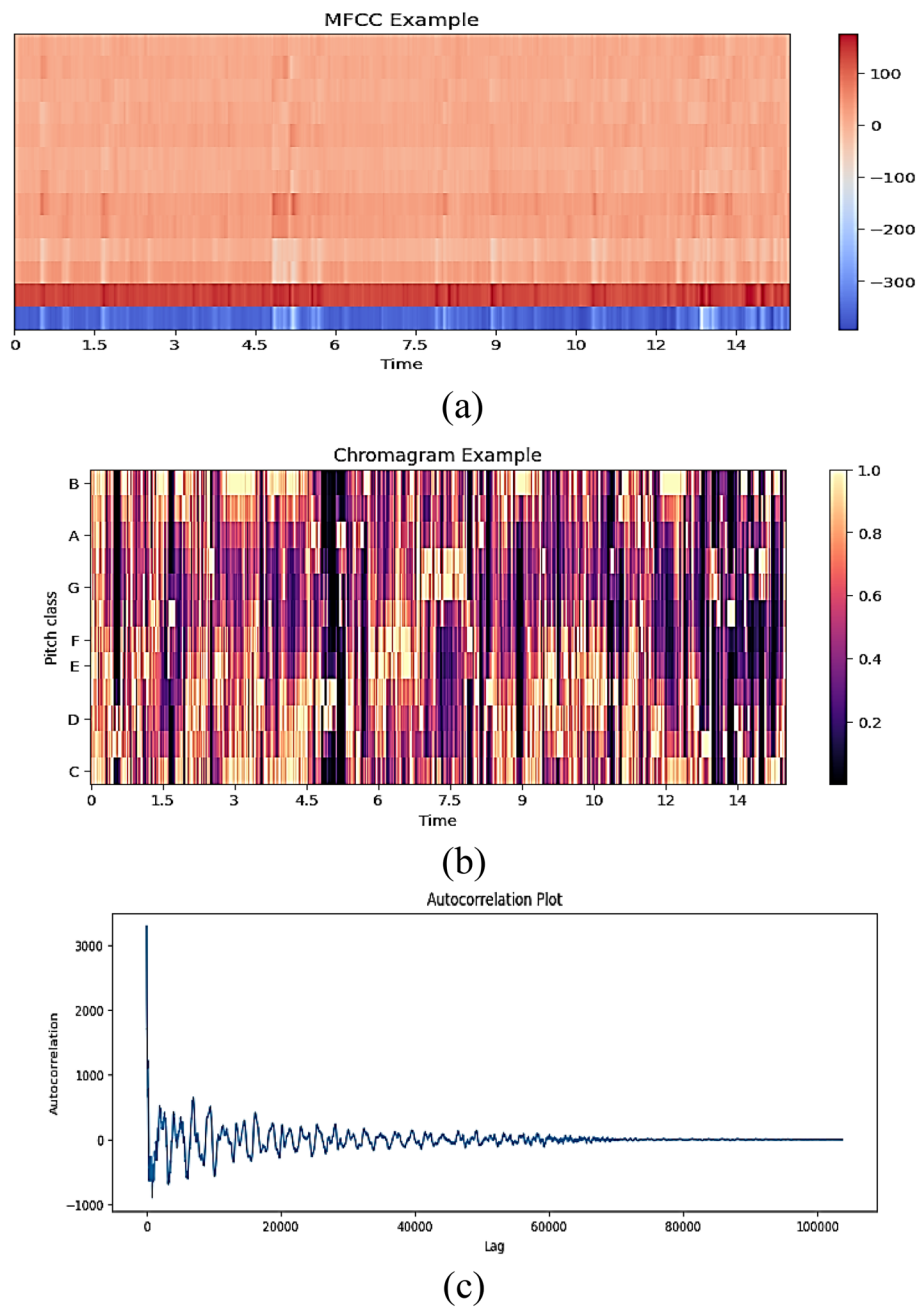
Third, our model is trained using the proposed algorithm which consists of using a Burn Layer with specified burn intensity and CNN with Adamax illustrated in Figs. 3 and 4. The proposed model resulting in MFCC features is displayed in Fig. 10a, while the chromogram illustrating the relationship between time and pitch size is depicted in Fig. 10b. Additionally, Fig. 10c showcases the autocorrelation plot, revealing lags ranging from 0 to 300,000. Subsequently, the learning curve of the proposed CNN + NADAM model is presented in Fig. 11.



**Fig. 6** The resulting **a** MFCC features, **b** chromogram, and **c** autocorrelation plot of the CNN and NADAM model

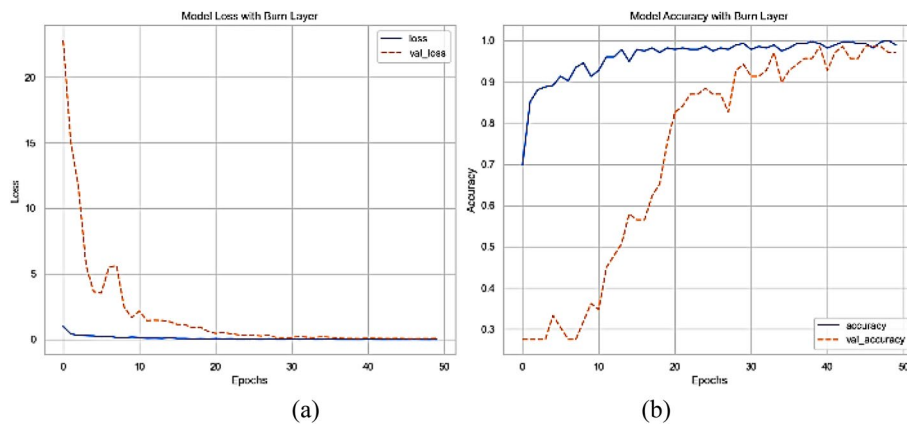


**Fig. 7** The Learning curve of the CNN + NADAM model **a** training and validation loss, **b** training and validation accuracy

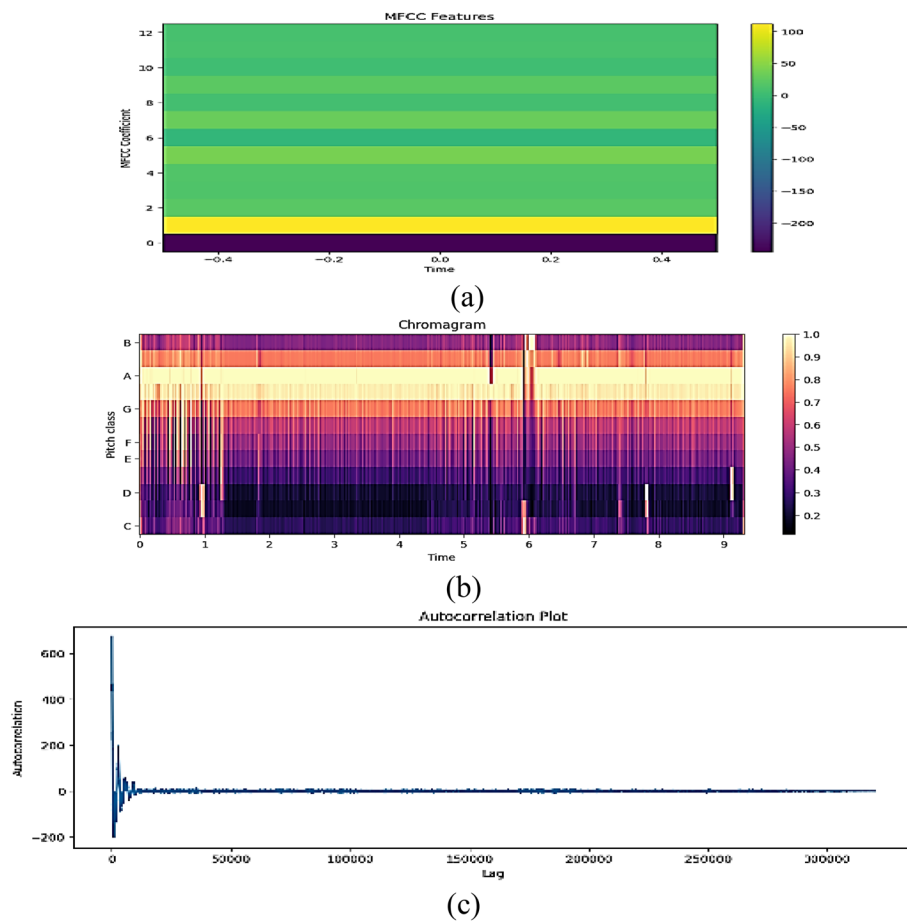


**Fig. 8** The resulting **a** MFCC features, **b** chromogram, and **c** autocorrelation plot of the ResNet (50) + Burn Layer + NADAM model

We performed experimental results using fivefold cross-validation along with standard deviation (SD) and the obtained results are shown in Table 5 as follows. The proposed model, evaluated using fivefold cross-validation, exhibits robust performance metrics with a sensitivity of  $89.96\% \pm 0.0528$ , specificity of  $82.58\% \pm 0.1124$ , precision of  $87.49\% \pm 0.0996$ , and accuracy of  $89.86\% \pm 0.0371$ . It also achieves a high F1 score of  $88.4\% \pm 0.0706$  and an MCC of  $72.00\% \pm 0.1260$ , demonstrating reliable and balanced classification capabilities as shown in Table 5.

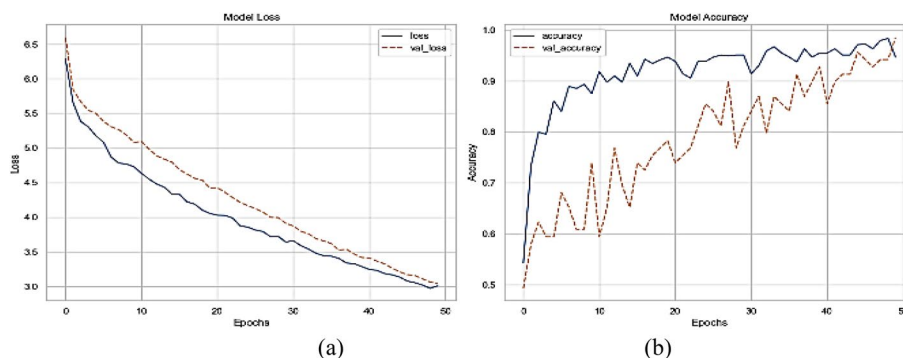


**Fig. 9** The Learning curve of the ResNet (50) + Burn Layer + NADAM model **a** training and validation loss, **b** training and validation accuracy



**Fig. 10** The resulting **a** MFCC features, **b** chromogram, and **c** autocorrelation plot of the proposed model

Table 6 showcases the performance comparison of three architectures: CNN + Nadam, ResNet (50) + Burn Layer + Nadam, the proposed model using fivefold cross-validation, and the proposed model based on Burn Layer with burn intensity parameter and



**Fig. 11** The Learning curve of the proposed model **a** training and validation loss, **b** training and validation accuracy

**Table 5** Performance metrics of the proposed model using fivefold cross validation

Architecture	Sensitivity (%) ± SD	Specificity (%) ± SD	Precision (%) ± SD	Negative predictive value (%) ± SD	Accuracy (%) ± SD	F1 score (%) ± SD	Matthews correlation coefficient (%) ± SD
The proposed model using fivefold cross-validation	89.96 ± 0.0528	82.58 ± 0.1124	87.49 ± 0.0996	84.35 ± 0.0704	89.86 ± 0.0371	88.4 ± 0.0706	72.00 ± 0.1260

**Table 6** The performance of three architectures CNN + Nadam, ResNet (50) + BurnLayer + Nadam, and the proposed model

Architecture	Sensitivity (%)	Specificity (%)	Precision (%)	Negative predictive value (%)	Accuracy (%)	F1 score (%)	Matthews correlation coefficient (%)
CNN + Nadam	89.29	86.96	89.29	86.96	88.24	89.29	76.24
ResNet (50) + Burn Layer + Nadam	97.56	93.75	93.02	97.83	95.51	95.24	91.08
The proposed model using fivefold cross-validation	89.96	82.58	87.49	84.35	89.86	88.40	72.00
The proposed model	96.77	100.00	100.00	95.00	98.55	98.36	95.88

Adamax optimizer. The evaluation is based on six standard classification metrics: sensitivity, specificity, precision, negative predictive value, accuracy, F1 score, and Matthews’s correlation coefficient.

As shown in Table 6, CNN + Nadam showed adequate yet modestly inferior results, achieving a sensitivity of 89.29% and an accuracy of 88.24%. ResNet (50) + Burn Layer + Nadam remarkably surpassed CNN + Nadam in specific areas, reaching a sensitivity of 97.56% and accuracy of 95.51%, though experiencing some decline in certain metrics. Nonetheless, the proposed work significantly outperformed the competition,



boasting a sensitivity of 96.77%, specificity of 100%, precision of 100%, negative predictive value of 95.00%, accuracy of 98.55%, F1 score of 98.36%, and Matthews’s correlation coefficient of 95.88%. These exceptional figures validate the proposed work’s prowess as a distinguished and competitive solution for poultry health status assessment tasks. The proposed model demonstrates 100% precision, reflecting its exceptional ability to classify positive cases accurately without any false positives. This performance was validated through fivefold cross-validation, revealing robust metrics: a sensitivity of 96.77%, specificity of 100%, and accuracy of 98.55%. These results highlight the model’s superior classification capabilities compared to previous methods, which achieved lower precision. The deep learning model integrates a custom Burn Layer that introduces random perturbations to the input data, enhancing the model’s resilience to varying signals. With a streamlined 19-layer architecture, including convolutional blocks, batch normalization, max pooling, and global average pooling, the model effectively handles audio classification tasks.

Table 7 presents the hyperparameters and their corresponding values used in the experimental setup. In determining the configuration of audio processing and training, hyperparameters play a crucial role in shaping the overall performance. In the audio analysis, Table 7 provides an exhaustive overview of these key parameters and their respective values employed in the experimental setup, pivotal for extracting meaningful features from audio signals and facilitating effective model training. The essential hyperparameters encompass the sampling frequency, where a rate of 44.1 kHz was utilized to ensure high-quality signal representation; the duration, specifying 2-s segments for analysis, allowing ample time for capturing pertinent audio information; the number of epochs, set at 50 to facilitate comprehensive learning from the data; the batch size, optimized at 32 to balance computational efficiency and model convergence; and the number of MFCC features, with 20 features chosen to capture relevant spectral information. These parameters collectively mold the audio processing pipeline and model training, enabling researchers to effectively analyze audio signals and extract valuable insights by selecting appropriate values.

To the best of our knowledge, since the dataset presented by Adebayo et al. in 2023 was published in the *Data in Brief Journal* [18], no other studies have cited or used this dataset. Therefore, we have made comparisons with related studies that used different datasets, as shown in Table 1. Additionally, in Table 8, we have focused on the audio files, which enable the diagnosis of poultry diseases.

The proposed model stands out by introducing a Burn Layer that injects controlled noise during training, enhancing robustness and generalization, a feature absent in

**Table 7** Hyperparameters and their corresponding values were used in the experimental setup

The hyperparameter labels	Value
Sampling frequency	44.1 kHz
Duration	2 s
Number of epochs	50
Batch size	32
Number of MFCC features	20

**Table 8** Comparative studies on diagnosing poultry diseases using audio files

Author	Methodology	Results
Huang et al. [43]	SVM	Accuracy = 90%
Carpentier et al. [42]	Deep learning models	Sensitivity = 88.40%
Quintana et al. [37]	DT	Accuracy = 86.10%
Cuan et al. [39]	ResNet-50 model	Accuracy = 91.06%
Xu and Chang [69]	YOLO V7 + LSTM	mAP = 86%
Proposed model	Accuracy = 98.55%	

other studies. It also develops an end-to-end pipeline specifically tailored for audio-based poultry health status detection, unlike most studies that rely on traditional machine learning or generic deep learning models. With an exceptional accuracy of 98.55%, the proposed model outperforms all compared methodologies. It efficiently reduces trainable parameters to 191,235, demonstrating high efficiency essential for practical deployment in resource-constrained environments. Comparative studies show other methodologies achieving lower results, such as Xu and Chang's YOLO V7 + LSTM with an mAP of 86%, Huang et al.'s SVM with 90% accuracy, Quintana et al.'s DT with 86.10% accuracy, Cuan et al.'s ResNet-50 model with 91.06% accuracy, and Carpentier et al.'s deep learning models with 88.40% sensitivity. The proposed model's contributions, including the Burn Layer and specialized pipeline, deliver markedly improved results in poultry disease diagnosis using audio files.

#### Statistical analysis: posthoc Nemenyi test

In this paper, we performed the statistical analysis using the Posthoc Nemenyi test which allows us to compare the pairs of models to determine which pairs are significantly different. The test produces a test statistic called the Nemenyi statistic, which is calculated as in Eq. (16).

$$Nemenyi\ statistic = \left(\frac{a}{b}\right)^2 - \left(\frac{c}{d}\right)^2 \quad (16)$$

where a and b are the accuracies of two models being compared, and c and d are the times required to achieve those accuracies. The p value for the Nemenyi test is calculated as in Eq. (17).

$$p\text{-value} = P(Nemenyi\ statistic > \text{observed Nemenyi statistic}) \quad (17)$$

The results of the posthoc Nemenyi test are as follows:

i) *CNN + Nadam vs. ResNet (50) + BurnLayer + Nadam*:

- Nemenyi statistic:  $\left(\frac{95.51}{88.24}\right)^2 - \left(\frac{1}{1}\right)^2 = 1.4014$
- p-value:  $P(Nemenyi\ statistic > 1.4014) = 0.1617$

ii) *CNN+Nadam vs. The Proposed Model:*

- Nemenyi statistic:  $\left(\frac{97.73}{88.24}\right)^2 - \left(\frac{1}{1}\right)^2 = 1.2352$
- p-value:  $P(\text{Nemenyi statistic} > 1.2352) = 0.2164$

iii) *ResNet (50) + BurnLayer + Nadam vs. The Proposed Model:*

- Nemenyi statistic:  $\left(\frac{97.73}{95.51}\right)^2 - \left(\frac{1}{1}\right)^2 = 0.0916$
- p-value:  $P(\text{Nemenyi statistic} > 0.0916) = 0.7626$

The p values from the Nemenyi test for all comparisons exceed the significance level of 0.05, indicating that the differences in accuracy between the models are not statistically significant. Although the proposed model demonstrates improvements in various performance metrics, these improvements are not statistically significant at the 0.05 level. Despite the lack of statistical significance, the proposed model exhibits the best overall performance: it achieves the highest accuracy (97.73%) and perfect sensitivity (100%), correctly identifying all positive cases. It also shows perfect specificity (100%), accurately identifying all negative cases. The model maintains high precision (95.00%) and the highest F1 score (95.88%), reflecting a strong balance between precision and recall. Additionally, with the highest Matthews Correlation Coefficient (95.17%), the proposed model demonstrates a robust correlation between observed and predicted classifications.

## Discussions and limitations

In this study, we propose a deep learning-based model for poultry audio classification, incorporating a Burn Layer for enhanced robustness. The model processes raw audio data and creates digital representations before applying a custom Burn Layer, which perturbs the input during training to improve model robustness. The architecture consists of convolutional blocks, global average pooling, fusion layers, and fully connected layers with another Burn Layer for additional robustness. Adamax optimizer is utilized to tackle the overfitting problem and improve the performance stability.

Compared to the traditional CNN+Nadam model, our proposed model demonstrated superior performance in terms of sensitivity, specificity, precision, negative predictive value, accuracy, F1 score, and Matthews's correlation coefficient. The inclusion of the Burn Layer proved advantageous in increasing model stability during training and exposed it to varying input, contributing to the model's overall efficacy.

Despite the promising results, limitations do exist in this study. First, the model's performance might degrade if confronted with extremely noisy or unstructured audio data since the Burn Layer introduces only controlled random noise. Further modifications may be needed to account for extreme cases. Second, expanding the dataset to cover a broader variety of poultry breeds and health conditions could strengthen the model's applicability and generalizability. Third, integrating transfer learning techniques could expedite model training and improve performance, particularly when limited training data is available.

Lastly, it is essential to acknowledge ethical concerns surrounding AI adoption in healthcare and veterinary settings. Ensuring privacy, fairness, and avoiding biases must

be priorities when leveraging AI for medical diagnoses. Transparent communication and collaboration among experts, policymakers, and stakeholders are crucial to establishing trustworthy and impactful AI solutions. Addressing these limitations and considerations will undoubtedly fuel ongoing research and drive advancements in deep learning-powered poultry health assessment.

### Conclusion and future work

This study presents a deep learning model for poultry audio classification that incorporates a custom Burn Layer to enhance robustness during training. The Burn Layer, which introduces random perturbations to input data, helps the model handle varying signals effectively. The model features a streamlined 19-layer architecture with three convolutional blocks, batch normalization, max pooling, and global average pooling, totaling 191,235 trainable parameters. Our model achieves impressive performance, with a sensitivity of 96.77%, specificity of 100.00%, and accuracy of 98.55%, surpassing previous methods in accuracy, precision, and recall.

Future research will focus on refining the Burn Layer by exploring adaptive burn intensities based on performance metrics, integrating recurrent networks to capture long-term dependencies, and validating the model with larger datasets. Additionally, investigating the effect of adversarial examples on model robustness and developing a real-time user interface for poultry health assessment are promising avenues. Our study utilized the dataset from Adebayo et al., which includes 346.wav files categorized as healthy (139), noise (86), and unhealthy (121). Future work will further define inclusion and exclusion criteria, particularly addressing challenges like overlapping voices, to improve dataset relevance and model accuracy.

### Author contributions

This work was carried out in collaboration among all authors. All Authors designed the study, performed the statistical analysis, and wrote the protocol. Authors MYS, TAEH, and EH managed the analyses of the study, managed the literature searches, and wrote the first draft of the manuscript. All authors read and approved the final manuscript.

### Funding

Open access funding provided by The Science, Technology & Innovation Funding Authority (STDF) in cooperation with The Egyptian Knowledge Bank (EKB).

### Data availability

The dataset used in this study is public and all test data are available at: <https://data.mendeley.com/datasets/zp4nf2dxbh/1>.

### Declarations

#### Consent for publication

This article does not contain any studies with human participants or animals performed by any of the authors.

#### Competing interests

The authors declare that they have no known competing interests.

Received: 19 March 2024 Accepted: 26 August 2024

Published online: 18 September 2024

### References

1. Jukan A, Masip-Bruin X, Amla N. Smart computing and sensing technologies for animal welfare: a systematic review. *ACM Comput Surv CSUR*. 2017;50(1):1–27.
2. Petso T, Jamisola RS Jr, Mpoeleng D. Review on methods used for wildlife species and individual identification. *Eur J Wildl Res*. 2022;68(1):3.

3. Vranken E, Mounir M, Norton T. Sound-based monitoring of livestock. In: Zhang Q, editor. *Encyclopedia of digital agricultural technologies*. Berlin: Springer; 2023. p. 1358–69.
4. Gibb R, Browning E, Glover-Kapfer P, Jones KE. Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring. *Methods Ecol Evol*. 2019;10(2):169–85.
5. Farrell DJ. Matching poultry production with available feed resources: issues and constraints. *Worlds Poult Sci J*. 2005;61(2):298–307.
6. Fontana I, Tullo E, Scrase A, Butterworth A. Vocalisation sound pattern identification in young broiler chickens. *Animal*. 2016;10(9):1567–74.
7. Laleye FA, Mousse MA. Attention-based recurrent neural network for automatic behavior laying hen recognition. *Multimed Tools Appl*. 2024;83:62443–58.
8. Tokuda I, Riede T, Neubauer J, Owren MJ, Herzel H. Nonlinear analysis of irregular animal vocalizations. *J Acoust Soc Am*. 2002;111(6):2908–19.
9. Tampuu A, Matisen T, Semikin M, Fishman D, Muhammad N. A survey of end-to-end driving: Architectures and training methods. *IEEE Trans Neural Netw Learn Syst*. 2020;33(4):1364–84.
10. Shams MY, Hassanien AE, Tang M. Deep belief neural networks for eye localization based speeded up robust features and local binary pattern. In: Shi X, Bohács G, Ma Y, Gong D, Shang X, editors. *LISS 2021. Lecture notes in operations research*. Singapore: Springer Nature; 2022. p. 415–30. [https://doi.org/10.1007/978-981-16-8656-6\\_38](https://doi.org/10.1007/978-981-16-8656-6_38).
11. Hassan E, Shams MY, Hikal NA, Elmougy S. The effect of choosing optimizer algorithms to improve computer vision tasks: a comparative study. *Multimed Tools Appl*. 2023;82(11):16591–633. <https://doi.org/10.1007/s11042-022-13820-0>.
12. Abdallah SE, Elmessery WM, Shams MY, Al-Sattary NSA, Abohany AA, Thabet M. Deep learning model based on ResNet-50 for beef quality classification. *Inf Sci Lett*. 2023;12(1):289–97.
13. Li Y, Chen Y, Wang N, Zhang Z. Scale-aware trident networks for object detection. In: *Proceedings of the IEEE/CVF international conference on computer vision*; 2019. p. 6054–63.
14. Salem H, Shams MY, Elzeki OM, Abd Elfattah M, Al-Amri JF, Elnazer S. Fine-tuning fuzzy KNN classifier based on uncertainty membership for the medical diagnosis of diabetes. *Appl Sci*. 2022;12(3):950.
15. Li X, et al. Efficient meta-tuning for content-aware neural video delivery. In: Avidan S, Brostow G, Cissé M, Farinella GM, Hassner T, editors., et al., *Computer vision—ECCV 2022. Lecture notes in computer science*. Cham: Springer Nature Switzerland; 2022. p. 308–24.
16. Shams MY, El-kenawy E-SM, Ibrahim A, Elshewey AM. A hybrid dipper throated optimization algorithm and particle swarm optimization (DTPSO) model for hepatocellular carcinoma (HCC) prediction. *Biomed Signal Process Control*. 2023;85:104908. <https://doi.org/10.1016/j.bspc.2023.104908>.
17. Abdelhamid AA, et al. Innovative feature selection method based on hybrid sine cosine and dipper throated optimization algorithms. *IEEE Access*. 2023;11:79750–76. <https://doi.org/10.1109/ACCESS.2023.3298955>.
18. Adebayo S, et al. Enhancing poultry health management through machine learning-based analysis of vocalization signals dataset. *Data Brief*. 2023;50:109528. <https://doi.org/10.1016/j.dib.2023.109528>.
19. Nam J, Choi K, Lee J, Chou S-Y, Yang Y-H. Deep learning for audio-based music classification and tagging: teaching computers to distinguish rock from bach. *IEEE Signal Process Mag*. 2018;36(1):41–51.
20. Li G, et al. Missing outcome data in recent perinatal and neonatal clinical trials. *Pediatrics*. 2024;153:e2023063101.
21. Morgan NK, Kim E, González-Ortiz G. Holo-analysis of the effects of xylo-oligosaccharides on broiler chicken performance. *Br Poult Sci*. 2024;65:79–86.
22. Nakrosis A, et al. Towards early poultry health prediction through non-invasive and computer vision-based dropping classification. *Animals*. 2023;13(19):3041. <https://doi.org/10.3390/ani13193041>.
23. He P, et al. Research progress in the early warning of chicken diseases by monitoring clinical symptoms. *Appl Sci*. 2022;12(11):5601. <https://doi.org/10.3390/app12115601>.
24. Mao Q, et al. Review detection of Newcastle disease virus. *Front Vet Sci*. 2022. <https://doi.org/10.3389/fvets.2022.936251>.
25. Machuve D, Nwankwo E, Mduma N, Mbelwa J. Poultry diseases diagnostics models using deep learning. *Front Artif Intell*. 2022. <https://doi.org/10.3389/frai.2022.733345>.
26. Liang J, Zhang C, Song J, Guo S. Research and prediction on initial contact pressure distribution of armature-rail contact surface under interference fit. 2024.p. 1–20.
27. Machuve D, Nwankwo E, Mduma N, Mbelwa J. Poultry diseases diagnostics models using deep learning. *Front Artif Intell*. 2022;5:733345. <https://doi.org/10.3389/frai.2022.733345>.
28. Cai Z, Cui J, Yuan H, Cheng M. Application and research progress of infrared thermography in temperature measurement of livestock and poultry animals: a review. *Comput Electron Agric*. 2023;205:107586. <https://doi.org/10.1016/j.compag.2022.107586>.
29. Caldara F, Nääs I, Garcia R. Infrared thermal image for assessing animal health and welfare. *J Anim Behav Biometeorol*. 2014;2:66–72. <https://doi.org/10.14269/2318-1265/jabb.v2n3p66-72>.
30. Yahav S, Giloh M. Infrared thermography—applications in poultry biological research. *Infrared Thermogr*. 2012. <https://doi.org/10.5772/27788>.
31. Nawaz AH, Amoah K, Leng QY, Zheng JH, Zhang WL, Zhang L. Poultry response to heat stress: its physiological, metabolic, and genetic implications on meat production and quality including strategies to improve broiler production in a warming world. *Front Vet Sci*. 2021;8:699081. <https://doi.org/10.3389/fvets.2021.699081>.
32. Noh J-Y, et al. Thermal image scanning for the early detection of fever induced by highly pathogenic avian influenza virus infection in chickens and ducks and its application in farms. *Front Vet Sci*. 2021;8:616755. <https://doi.org/10.3389/fvets.2021.616755>.
33. Chuang C-H, Chiang C-Y, Chen Y-C, Lin C-Y, Tsai Y-C. Goose surface temperature monitoring system based on deep learning using visible and infrared thermal image integration. *IEEE Access*. 2021;9:131203–13. <https://doi.org/10.1109/ACCESS.2021.3113509>.

34. Gourisaria MK, Arora A, Bilgaiyan S, Sahni M. Chicken disease multiclass classification using deep learning, vol. 614 LNNS. Singapore: Springer Nature Singapore; 2023.
35. Carroll B, Anderson D, Daley W, Harbert S, Britton D, Jackwood M. Detecting symptoms of diseases in poultry through audio signal processing. In: IEEE global conference on signal and information processing, Global 2014. 2015. p. 1132–5. <https://doi.org/10.1109/GlobalSIP.2014.7032298>.
36. Aydin A, Berckmans D. Using sound technology to automatically detect the short-term feeding behaviours of broiler chickens. *Comput Electron Agric.* 2016;121:25–31. <https://doi.org/10.1016/j.compag.2015.11.010>.
37. Quintana MMD, Infante RRD, Torrano JCS, Pacis MC. A hybrid solar powered chicken disease monitoring system using decision tree models with visual and acoustic imagery. In: 2022 14th International conference on computer and automation engineering ICCAE; 2022. p. 65–9.
38. Li Z, et al. Sex detection of chicks based on audio technology and deep learning methods. *Anim Open Access J MDPI.* 2022;12(22):3106. <https://doi.org/10.3390/ani12223106>.
39. Cuan K, Zhang T, Li Z, Huang J, Ding Y, Fang C. Automatic Newcastle disease detection using sound technology and deep learning method. *Comput Electron Agric.* 2022;194(January):106740. <https://doi.org/10.1016/j.compag.2022.106740>.
40. Jakovljević N, Maljković N, Mišvsković D, Knežević P, Delić V. A broiler stress detection system based on audio signal processing. In: 2019 27th telecommunication forum TELFOR; 2019. p. 1–4.
41. Wang C, Benetos E, Wang S, Versace E. Joint scattering for automatic chick call recognition. In: 2022 30th European signal processing conference (EUSIPCO); 2022. p. 195–9. <https://doi.org/10.23919/EUSIPCO55093.2022.9909738>.
42. Carpentier L, Vranken E, Berckmans D, Paeshuyse J, Norton T. Development of sound-based poultry health monitoring tool for automated sneeze detection. *Comput Electron Agric.* 2019;162:573–81.
43. Huang J, Wang W, Zhang T. Method for detecting avian influenza disease of chickens based on sound analysis. *Biosyst Eng.* 2019;180:16–24. <https://doi.org/10.1016/j.biosystemseng.2019.01.015>.
44. Jamshidi H, Budak E. On the prediction of surface burn and its thickness in grinding processes. *CIRP Ann.* 2021;70(1):285–8.
45. Suha SA, Sanam TF. A deep convolutional neural network-based approach for detecting burn severity from skin burn images. *Mach Learn Appl.* 2022;9:100371.
46. Zhang P, Nascetti A, Ban Y, Gong M. An implicit radar convolutional burn index for burnt area mapping with Sentinel-1 C-band SAR data. *ISPRS J Photogramm Remote Sens.* 2019;158:50–62.
47. Shrivastava AK, Sharma A, Awale AS, Yusufzai MZK, Vashista M. Assessment of grinding burn of AISI D2 tool steel using Barkhausen noise technique. *J Inst Eng India Ser C.* 2021;102(4):885–96.
48. Jiang S, Wang Y, Wang Y. SelfEvolve: a code evolution framework via large language models. *ArXiv Preprint arXiv:2306.02907*, 2023.
49. Cirillo MD, Mirdell R, Sjöberg F, Pham TD. Time-independent prediction of burn depth using deep convolutional neural networks. *J Burn Care Res.* 2019;40(6):857–63. <https://doi.org/10.1093/jbcr/irz103>.
50. Salehin I, Kang D-K. A review on dropout regularization approaches for deep neural networks within the scholarly domain. *Electronics.* 2023;12(14):3106. <https://doi.org/10.3390/electronics12143106>.
51. Liu T, Chen T, Niu R, Plaza A. Landslide detection mapping employing CNN, ResNet, and DenseNet in the Three Gorges reservoir, China. *IEEE J Sel Top Appl Earth Obs Remote Sens.* 2021;14:11417–28.
52. Weng O, et al. Tailor: altering skip connections for resource-efficient inference. *ACM Trans Reconfig Technol Syst.* 2024;17(1):1–23.
53. Yuan X, Savarese P, Maire M. Accelerated training via incrementally growing neural networks using variance transfer and learning rate adaptation. *Adv Neural Info Process Syst.* 2024;36:16673–16692.
54. Thakur N, Bhattacharjee E, Jain R, Acharya B, Hu Y-C. Deep learning-based parking occupancy detection framework using ResNet and VGG-16. *Multimed Tools Appl.* 2024;83(1):1941–64.
55. Hu Y, Deng L, Wu Y, Yao M, Li G. Advancing spiking neural networks toward deep residual learning. In: *IEEE transactions on neural networks and learning systems.* 2024. pp 1–15.
56. Hassan E, Hossain MS, Saber A, Elmougy S, Ghoneim A, Muhammad G. A quantum convolutional network and ResNet (50)-based classification architecture for the MNIST medical dataset. *Biomed Signal Process Control.* 2024;87:105560. <https://doi.org/10.1016/j.bspc.2023.105560>.
57. Antonio CB, Bautista LGC, Labao AB, Naval PC. Vertebra fracture classification from 3D CT lumbar spine segmentation masks using a convolutional neural network. In: Nguyen NT, Hoang DH, Hong T-P, Pham H, Trawiński B, editors. *Intelligent information and database systems Lecture notes in computer science.* Cham: Springer International Publishing; 2018. p. 449–58. [https://doi.org/10.1007/978-3-319-75420-8\\_43](https://doi.org/10.1007/978-3-319-75420-8_43).
58. Alom MZ, Hasan M, Yakopcic C, Taha TM, Asari VK. Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation. *arXiv preprint arXiv:1802.06955*, 2018.
59. Sharan RV, Moir TJ. An overview of applications and advancements in automatic sound recognition. *Neurocomputing.* 2016;200:22–34.
60. Incze A, Jancsó H-B, Szilágyi Z, Farkas A, Sulyok C. Bird sound recognition using a convolutional neural network. In: 2018 IEEE 16th international symposium on intelligent systems and informatics (SISY). IEEE; 2018. p. 000295–300.
61. Zhang H, McLoughlin I, Song Y. Robust sound event recognition using convolutional neural networks. In: 2015 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE; 2015. p. 559–63.
62. Mehta S, Paurwala C, Vaidya B. CNN based traffic sign classification using Adam optimizer. In: 2019 international conference on intelligent computing and control systems (ICCS). IEEE; 2019. p. 1293–8.
63. Vani S, Rao TM. An experimental approach towards the performance assessment of various optimizers on convolutional neural network. In 2019 3rd international conference on trends in electronics and informatics (ICOEI). IEEE; 2019. p. 331–6.
64. Yaqub M, et al. State-of-the-art CNN optimizer for brain tumor segmentation in magnetic resonance images. *Brain Sci.* 2020;10(7):427.

65. Kumar A, Sarkar A, Pradhan C. Malaria disease detection using CNN technique with SGD, RMSprop and ADAM optimizers. In: Dash S, Acharya B, Mittal M, et al. (eds) Deep learning techniques for biomedical and health informatics. Studies in Big Data, vol. 68. Cham: Springer, 2020. pp. 211–230.
66. Shams MY, Abd El-Hafeez T, Hassan E. Acoustic data detection in large-scale emergency vehicle sirens and road noise dataset. *Expert Syst Appl*. 2024;249:123608. <https://doi.org/10.1016/j.eswa.2024.123608>.
67. Hassan E, Shams MY, Hikal NA, Elmougy S. A novel convolutional neural network model for malaria cell images classification. *Comput Mater Contin*. 2022;72(3):5889–907. <https://doi.org/10.32604/cmcc.2022.025629>.
68. Sarhan S, Nasr AA, Shams MY. Multipose face recognition-based combined adaptive deep learning vector quantization. *Comput Intell Neurosci*. 2020;2020:1–11.
69. Xu R-Y, Chang C-L. Deep learning-based poultry health diagnosis: detecting abnormal feces and analyzing vocalizations. In: 2024 10th international conference on applied system innovation (ICASI). 2024. p. 55–7. <https://doi.org/10.1109/ICASI60819.2024.10547723>.

### **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.