Gear Defect Detection Using Hybrid CNN-LSTM

Sanjivani Atakare, Prof. A. M. Dyade
SVERI's College Of Engineering Pandharpur, Maharashtra, India

Abstract: -

Gear assemblies are essential components in many industrial machines, and even minor defects can lead to unplanned shutdowns, financial losses, and possible safety concerns. Detecting faults at an early stage is therefore crucial for predictive maintenance. This study introduces a hybrid deep learning model that combines Convolutional Neural Networks (CNN) with Long Short-Term Memory (LSTM) units to automatically diagnose gear faults using vibration data. In this method, the collected vibration signals are first transformed into spectrograms so that important frequency-based information becomes more visible. The CNN layers learn spatial features from these spectrograms, while the LSTM layers capture the time-dependent patterns that reflect different gear conditions. By integrating these two architectures, the proposed system delivers more reliable and accurate fault detection than traditional machine learning approaches or standalone CNN/LSTM models. Tests conducted on a standard gear-fault dataset—including conditions like gear cracks, tooth wear, and broken teeth—show that the hybrid CNN–LSTM framework provides superior accuracy and greater robustness. This makes it a strong candidate for real-time, automated monitoring of gear health in industrial settings.

Keywords: Gear Defect Detection, CNN-LSTM Hybrid Model, Condition Monitoring, Vibration Analysis, Deep Learning, Predictive Maintenance, Fault Diagnosis, Time-Frequency Analysis, Spectrogram, Industrial Automation.

1. Introduction

Gears are essential elements in many mechanical applications, ranging from vehicle transmission systems to heavy industrial equipment. Their main function is to transfer motion and power smoothly and efficiently. Because they operate continuously under varying loads, gears often develop faults such as wear, cracking, pitting, or tooth breakage. If these issues are not identified in time, they can cause serious failures, unexpected shutdowns, and high maintenance expenses. This makes early and precise fault detection extremely important for maintaining reliability, safety, and overall performance of mechanical systems.Conventional approaches to gear fault diagnosis depend largely on manual inspection, handcrafted feature extraction, and traditional signal-processing methods. These techniques are labor-intensive, require considerable expertise, and usually do not adapt well to changing operating conditions. With the rapid progress in AI and DL, automated fault detection has become a strong alternative because it can learn complex patterns directly from raw or lightly processed vibration data.CNNs have shown exceptional capability in extracting spatial information from images and spectrograms produced from vibration signals. However, they are not ideal for capturing the temporal relationships in time-series data. In contrast, recurrent neural networks, especially LSTM models, are effective at learning sequential dependencies but may not efficiently represent localized spatial details.To address these limitations, the present study introduces a hybrid DL model that integrates both CNN and LSTM architectures. In this framework, vibration signals collected from gear systems are converted into spectrograms using time–frequency analysis. The CNN layers extract key spatial characteristics from these spectrograms, while the LSTM layers learn the temporal behavior associated with different gear conditions. Combining strengths of both models, the hybrid CNN LSTM approach enhances accuracy and improves robustness in classifying various gear defects.The model is tested on standard gear-fault datasets containing several defect types under diverse operating conditions. Experimental findings indicate that the CNN–LSTM hybrid system performs better than traditional ML methods and individual DL models, making it a strong candidate for real time gear health monitoring in industrial settings.Table 1 summarizes conventional techniques along with their limitations.

Table1 Traditional Techniques Used for gear defect detection and Limitation

Sr. No.	Traditional Technique Used	Application in Gear Defect Detection	Limitations of Technique
1	Visual Inspection	Detecting surface defects like cracks, pitting, and wear	Subjective, labor-intensive, not reliable for internal defects, prone to human error
2	Vibration Analysis (Classical FFT)	Monitoring gear mesh frequency changes indicating defects	Requires manual feature extraction, limited sensitivity to early-stage faults
3	Acoustic Emission Analysis	Identifying crack initiation and material deformation in gears	Sensitive to noise, complex signal interpretation, requires expertise
4	Oil Debris Analysis	Detecting wear by analyzing metal particles in lubricants	Cannot pinpoint defect location or type, indirect indication only
5	Thermography	Detecting temperature changes due to friction from gear misalignment/wear	Affected by ambient conditions, low resolution, unsuitable for internal defect identification
6	MPI	Detecting surface and near surface cracks in ferromagnetic gears	Not suitable for non-ferromagnetic materials, time-consuming, requires gear disassembly
7	Ultrasonic Testing	Locating internal flaws and cracks in gears	Requires skilled operator, not effective on complex geometries, slow for high-volume testing
8	Strain Gauge Analysis	Monitoring gear load and deformation behavior under stress	Requires physical sensor attachment, affected by noise, not ideal for real-time fault detection

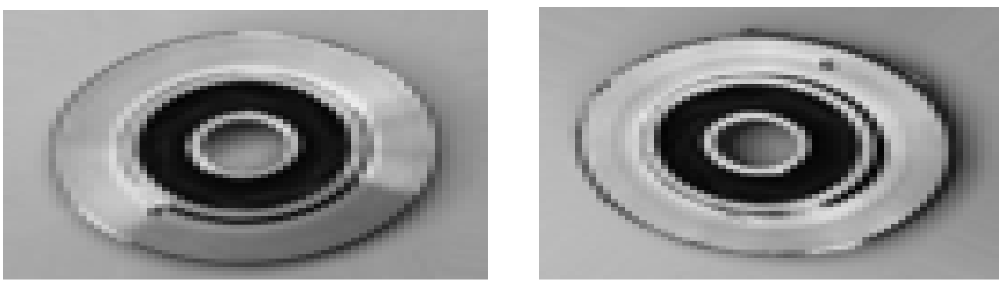

Fig. 1. Images captured across various defects

The key contribution of the work as follows

- Development of a hybrid CNN-LSTM model specifically designed for gear defect detection using vibration and acoustic data.
- Transformation of raw signal data into spectrograms for effective spatial and temporal feature extraction.
- Construction and preprocessing of a comprehensive gear fault dataset to train and evaluate the model under varying defect conditions.
- Extensive performance evaluation comparing hybrid model with standalone CNN and LSTM architectures to highlight its improved accuracy and robustness.
- Demonstration of model's feasibility for real time, automated fault detection in industrial gear systems, contributing to predictive maintenance and reduced downtime.

This research bridges gap between traditional fault diagnosis techniques and modern DL approaches, advancing field of intelligent condition monitoring in mechanical systems.

Literature Survey

Paper introduces a DL framework for the identification and classification of surface defects, using CNNs. Important feature of research is creation of artificial data by means of GANs to enlarge the training datasets, thus improving accuracy of the model in case of scarce real-world data.The method addresses challenges in surface defect detection in manufacturing processes, where defects can be subtle and varied. The CNN model is trained on both real and synthetic images, significantly improving its ability to classify and detect surface anomalies. The paper presents a thorough evaluation of the model, comparing it against traditional defect detection methods. It suggests that this approach is highly scalable and can be integrated into automated quality control systems. The findings offer valuable insights into use of DL for industrial applications. Finally, the study concludes with recommendations on future research directions, including improving synthetic data quality [7].

This author discusses applying CNN and LSTM networks to predict melt pool temperatures during wire arc additive manufacturing (WAAM). The approach combines CNNs feature extraction and LSTMs for temporal prediction, making it suitable for dynamic processes. The CNN is responsible for capturing spatial patterns from images, while LSTM captures temporal dependencies in data. Model successfully predicts thermal behavior, which is critical for optimizing the additive manufacturing process. The authors show that this hybrid deep learning approach outperforms traditional methods for terms of accuracy. Their results demonstrate that ML can enhance precision of temperature predictions, leading to better control over the WAAM process. The paper highlights the importance of real-time prediction for improving the quality of printed parts. Furthermore, authors discuss potential for integrating this model into an industrial setting for automated quality control. The paper concludes that combining CNNs and LSTMs offers significant advantages over single-model approaches in complex manufacturing tasks. Future work on optimizing model further and incorporating additional features for improved predictions [8].

This paper author presents a tool wear prediction model based on combination of CNNs and bidirectional LSTMs, enhanced by Improved Particle Swarm Optimization (IPSO). The model is designed to analyze cutting force, vibration, and acoustic emission signals, which are key indicators of tool wear in manufacturing. The use of bidirectional LSTMs allows model to capture both past and future temporal dependencies, improving prediction accuracy. Integration of CNNs helps model Learns useful patterns directly from raw data, reducing reliance on manually crafted feature. IPSO applied to optimize model parameters, ensuring that the prediction is as accurate as possible. The study demonstrates that this hybrid model outperforms traditional machine learning approaches in terms of prediction precision. Results show that model can effectively predict tool wear, leading to optimized maintenance schedules and reduced downtime [9].

The system focuses on using deep learning, specifically CNNs, for detecting and classifying surface defects in steel products. The authors demonstrate how CNNs can automatically detect various types of defects, including scratches, pits, and cracks, which are commonly found during the manufacturing process. Model is trained on large dataset of steel surface images, where it learns distinguish between defective and non-defective regions. One key challenges addressed in the study is variance in defect types and sizes, which makes traditional methods less effective. The authors use a data augmentation technique to improve model's ability to generalize, thereby enhancing its robustness to different defect patterns. Results show that DL approach achieves high accuracy, significantly reducing time and labor required for manual inspection. Model is evaluated on real-world data, demonstrating its feasibility for industrial applications. The paper discusses the potential for integrating the model into automated quality control systems, offering a promising solution for improving manufacturing efficiency [10].

This explores the evolution of object detection techniques, focusing on the integration of CNNs and Vision Transformers (ViTs). The authors provide a comprehensive overview of the history and development of CNNs, highlighting their significant impact on object detection tasks. They then discuss the emergence of Vision Transformers, which have demonstrated superior performance in several domains. The paper compares CNNs and ViTs in terms of architecture, efficiency, and accuracy, explaining why ViTs are gaining popularity. One of the key points discussed is the ability of ViTs to capture global contextual information, which gives them an edge in complex detection scenarios [11].

This paper proposes an intrusion detection system for IoT networks, integrating CNN-LSTM networks with statistical filtering techniques. The CNNs are used to extract relevant features from IoT traffic data, while LSTM network captures temporal dependencies in sequence of network activities. The statistical filtering technique helps in reducing false alarms by filtering out irrelevant data and focusing on critical events. The model is evaluated on dataset of IoT traffic and demonstrates superior performance in detecting both known and unknown intrusion patterns. The hybrid CNN-LSTM approach is shown to provide higher accuracy than traditional ML models.Authors highlight the importance of temporal analysis in intrusion detection, especially in IoT environments where the data streams are continuous and dynamic. The paper discusses the challenges of real-time intrusion detection and the need for efficient processing techniques. The model is designed to be scalable, making it suitable for large IoT networks. In conclusion, the paper suggests that combining deep learning with statistical methods can significantly improve accuracy and reliability of intrusion detection systems [12].

Author presented on the automatic detection and quantification of defects on hot-rolled steel surfaces using deep learning. The authors employ CNNs to process high-resolution images of steel surfaces and identify defects such as cracks, scale, and pitting. The CNN model is trained to classify defects into different categories, providing a robust solution for quality inspection in the steel industry. The paper introduces a pixel-level segmentation technique, which enables precise localization and quantification of defects. This approach not only detects the presence of defects but also estimates their size and severity, offering detailed feedback for quality control. Model evaluated is on large dataset of steel surface images, demonstrating its accuracy and reliability. The results show that CNN model significantly outperforms traditional image processing techniques in terms of defect detection accuracy. Paper discusses potential for integrating this deep learning approach into industrial production lines, where it could operate in real-time [13].

Faulty gears often produce loud, irregular, and highly non-stationary noise during operation. Severe defects such as broken teeth can further damage the entire gear system if not detected early. This work introduces a diagnostic approach based on the dual tree complex wavelet transform for identifying gear faults. Acoustic signal from a healthy gear mesh is used as reference for comparison. Experiments were performed on gears with one or more teeth containing embedded defects. The method is capable of locating the angular positions of multiple damaged teeth with high precision.

Proposed method is employed for detection performance by means of running simulations. Parameters such as like mean, auto correlation, dynamic range, standard deviation, and crest factor are measured. Acoustic signals are used for gear fault identification as well as defect severity estimation. It is instrumental in the prevention of gear tooth fractures and the motors gear health monitoring [14].

Synthetic bevel gear inspection system using a multi camera vision setup has been developed to address these challenges. The machine is designed to evaluate gear dimensions and detect surface defects at the same time. It employs three environmentally friendly algorithms—NAD, CAM, and FRP—within its processing pipeline. The system can identify a wide range of imperfections, including dents, scratches, porosity, cracks, impacts, and repeated spline damage. It offers dimensional accuracy in the range of 40–50 µm, with the smallest detectable defect measuring about 0.4 mm. Each inspection takes only around 1.3 seconds, providing sufficient speed and precision for real-time, in-line quality monitoring during bevel-gear production. [15]

In the early stages of equipment degradation, small local micro-fractures can generate distinct magnetic flux leakage (MFL) signals, as explained by magnetic dipole–based MFL theory. The characteristics of a crack are reflected in the MFL waveform through features such as its height along the tangential direction, the peak value of the gradient, and the zero-crossing points around the defect zone. These findings form the theoretical foundation for the metal magnetic memory (MMM) method. Using this principle, micro-cracks on actual equipment were identified through both dynamic loading tests and static measurements. The influence of operational loads on detection accuracy was also examined. Results show that the gradient peak in the MFL signal provides a reliable indication of defect presence and position, confirming the effectiveness of magnetic memory testing for early fault identification in mechanical components. [16].

Railway gear is an essential part. The gear's mention of being in a bad state is what safety and quality of train travel are mainly affected. Paper proposes an automatic and quantitative method for disordered state determination of external tools. A degree method is first introduced for segmentation of meshing position in equipment teeth. Then, to identify floor defects, adaptive threshold, and formative evaluation are combined. Strategy for detecting illnesses is overall more efficient than some of the related methods [17].

Material and methods

Dataset and key consideration

Dataset used in this study contains vibration and acoustic signals collected from industrial gearboxes operating under varying load and speed conditions. Signals were recorded using high-precision accelerometers and microphones at a sampling rate between 12--20 kHz to capture gear mesh frequencies and fault-related harmonics. The dataset includes six gear health conditions: healthy gear, tooth wear, crack, pitting, broken tooth, and misalignment. Each condition contains multiple time–series segments, which were further converted into spectrograms for CNN based spatial feature extraction. All samples were labeled according to the corresponding fault type and split into three parts: 70% for training, 15% for validation, and 15% for testing.

Publicly available benchmark vibration datasets were also incorporated to increase variability and strengthen model robustness. Final dataset provides diverse and representative patterns essential for accurate gear defect detection.

Signal Processing

Signal processing is critical step in automated detection and classification of gear defects. It transforms raw vibration or acoustic signals into meaningful representations suitable for DL. Process begins with segmentation of the time-series data into fixed-length windows to maintain consistency across input samples. Noise reduction is then applied using techniques such as low-pass or band-pass filtering to eliminate high-frequency noise and unwanted signal components. These steps ensure that only relevant frequency and temporal features are retained for analysis. The cleaned signals are then converted into 2D time-frequency representations like STFT spectrograms or Continuous Wavelet Transforms making them suitable for CNN input layers. Normalization is applied to the spectrograms to scale values uniformly, which helps the network converge faster and prevents bias toward specific amplitude ranges. These preprocessing steps ensure input data is clean, consistent, highly representative of actual gear defect characteristics.

Data Augmentation

Data augmentation in gear defect detection aims improve model robustness and generalization by introducing variability into the training data. Given that collecting large-scale vibration datasets in real-world industrial settings can be challenging, augmentation techniques help simulate diverse operating conditions. Common augmentation techniques include:

- **Adding Gaussian noise** to simulate sensor variability,
- **Time shifting** to account for phase differences,
- **Time stretching or compression** to mimic speed variations,
- **Amplitude scaling** to simulate signal gain differences,
- **Frequency masking or shifting** to alter dominant fault frequencies.

By generating multiple variations of the same signal or spectrogram, the CNN-LSTM model becomes better equipped to identify gear defects under different operating loads, rotational speeds, and noise levels—making the system more accurate and reliable in real-time industrial applications.

Proposed System: -

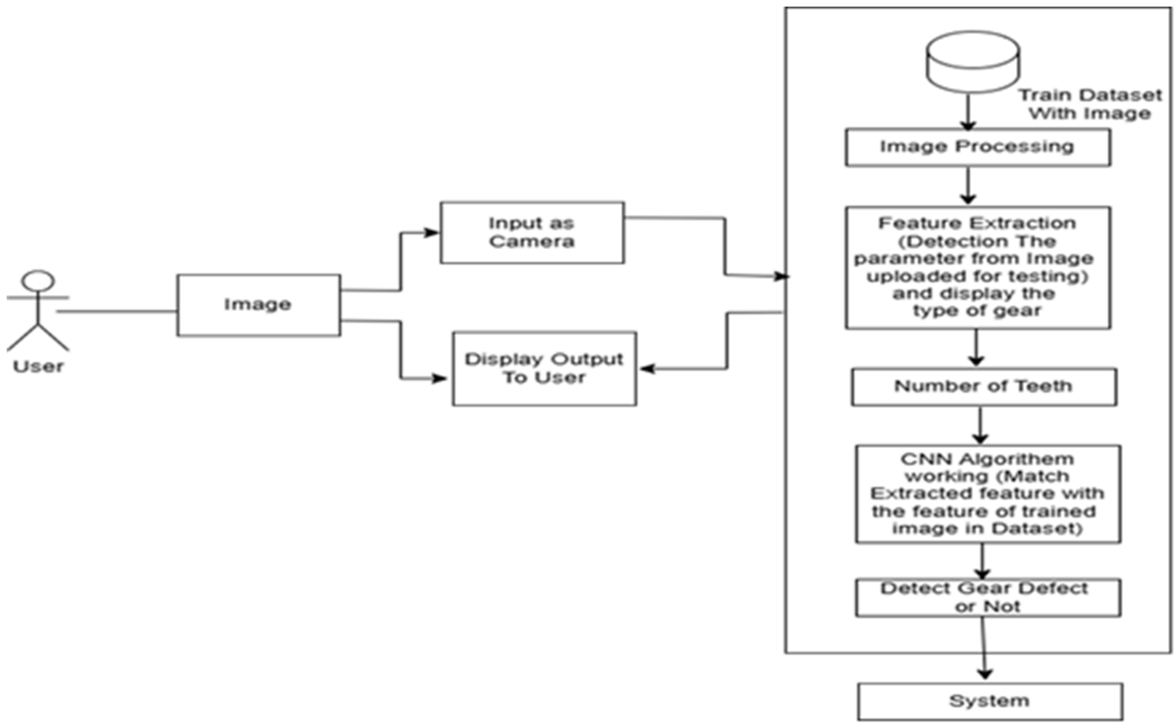

Fig 2 Proposed System Architecture

Hybrid CNN-LSTM model is powerful DL architecture designed for enhance accuracy of gear defect detection by leveraging both spatial and temporal feature learning. In model CNNs are first employed to extract essential spatial features from time–frequency representations of vibration or acoustic signals, such as spectrograms. These features include harmonic patterns, frequency shifts, amplitude variations, and other signatures associated with gear defects such as wear, cracks, pitting, or misalignment. The CNN layers—comprising convolutional, activation (ReLU), and pooling layers—automatically learn these discriminative patterns while reducing data dimensionality without losing important diagnostic information.

Rescaling Layer

The model first normalizes the raw images through a rescaling step, converting all pixel values into the range [0, 1]. This helps maintain consistency across inputs and improves training stability by keeping gradients well-behaved during backpropagation.

Data Augmentation Layer

To enhance generalization and reduce overfitting, the images undergo random alterations such as contrast changes, flipping, and rotation, zooming, and shifting. These transformations generate diverse versions of the original images, enriching the training set.

First Convolutional Layer

This layer applies 32 filters size 3×3 with *same* padding to learn basic visual cues, such as edges and simple shapes. ReLU activation introduces non linearity, enabling model to capture more complex patterns.

First Max-Pooling Layer

A 2×2 max-pooling operation reduces spatial dimensions of feature maps, retaining only most relevant details. This step lowers computational demand while preserving dominant features.

Second Convolutional Layer

Using 64 filters of size 3×3, this layer identifies more detailed and deeper visual patterns based on the representations extracted earlier. ReLU activation is again applied to support non-linear feature learning.

Second Max-Pooling Layer

Another 2×2 pooling layer further compresses the feature maps, highlighting essential information and reducing redundancy.

Third Convolutional Layer

With 128 filters, this layer expands feature depth and extracts higher-level characteristics such as texture, shape variations, and other complex structures. It plays a crucial role in advanced feature abstraction.

Third Max-Pooling Layer

This pooling step continues to downsample the feature maps, shrinking dimensionality while keeping the most influential visual patterns. This helps prepare the data for sequence-based processing.

Reshape Layer

Output of final pooling layer is reshaped into sequence-like format, converting 2D feature maps into 3D structures suitable for input into the LSTM. This transformation creates a temporal interpretation of the spatial information.

LSTM Layer

LSTM layer with 128 units learns relationships across the sequentially arranged feature data. It captures long-range dependencies and helps the model identify fault patterns that span different regions of image.

Fully Connected Dense Layers

The LSTM output is fed into two dense layers 512 and 256 neurons, each followed by dropout rate of 50%. These layers refine the learned representation and aid in forming stronger classification boundaries.

Output Layer

The final softmax layer produces probability scores for each fault category. It converts extracted features into predicted class, indicating most likely gear-fault type present in the input image.

Table 2 Hyper Parameter sets

Hyperparameter	Value
Input Size	(224, 224, 3)
Conv Filters	[16, 32, 64]
Kernel Size	(3, 3)
Activation	ReLU
Pooling Size	(2, 2)
Dropout Rate	0.5
Dense Units	128
LSTM Units	64
Loss Function	categorical_crossentropy
Optimizer	Adam (learning_rate=0.001)

1. Rescaling Layer

Equation:

$x\_normalized = x / 255$……………………………………………………….eq (1)

Purpose: Normalizes pixel values to a range [0, 1], making model training more stable and faster.

2. Convolutional Layer

Equation:

$F(i, j) = \Sigma \Sigma K(m, n) \cdot I(i + m, j + n)$…………………………………………..eq(2)

Purpose: Applies a kernel K over input image I to extract spatial features like edges and textures.

3. ReLU Activation

Equation:

$ReLU(x) = \max(0, x)$ ………………………………………………………….eq(3)

Purpose: Introduces non linearity in model, allowing it learn complex patterns.

4. Max Pooling Layer

Equation:

$P(i, j) = \max\{x \in pool(i, j)\}$…………………………………………………….eq(4)

Purpose: Down samples feature map by selecting maximum value in each pool window, reducing dimensionality.

5. Flatten Layer

Equation:

Flatten: $R^{h \times w \times d} \rightarrow R^{h \cdot w \cdot d}$…………………………………………….eq(5)

Purpose: Converts multi-dimensional output to one-dimensional vector for input into dense layers.

6. Dense (Fully Connected) Layer

Equation:

y = f(Wx + b) ……………………………………………………………………eq(6)

Purpose: Combines all input features to produce the final prediction. f is typically ReLU or Softmax.

7. LSTM Layer (with gates)

Equation:

Forget Gate: ft = σ(Wf · [ht-1, xt] + bf)

Input Gate: it = σ(Wi · [ht-1, xt] + bi)

Candidate Cell: C~t = tanh(WC · [ht-1, xt] + bC)

Cell Update: Ct = ft * Ct-1 + it * C~t

Output Gate: ot = σ(Wo · [ht-1, xt] + bo)

Hidden State: ht = ot * tanh(Ct)

Purpose: LSTM layer learns time based patterns. Each gate controls what information to keep or discard.

8. Softmax Output Layer

Equation:

Softmax (zi) = e^{zi} / Σ e^{zj}…………………………………………………eq(7)

Purpose: Converts final outputs into probabilities for multi-class classification tasks.

Mathematical Formulation

Gear defect detection begins with a vibration or acoustic signal$\boldsymbol{s}(\boldsymbol{t})$, which is divided into fixed-length segments$x_1, x_2, \ldots, x_n$. Each segment is transformed into a timefrequency representation using STFT:

$$S_i(\tau, f) = |\mathrm{STFT}\{x_i(t)\}|^2,$$

Producing a spectrogram used as the CNN input. The CNN extracts spatial feature vectors

$$\mathbf{z}_i = \phi_{\mathrm{CNN}}(S_i),$$

Which are then fed sequentially into LSTM to model temporal dependencies:

$$\mathbf{h}_t = \mathrm{LSTM}(\mathbf{z}_t, \mathbf{h}_{t-1}).$$

The final hidden state $\mathbf{h}_n$ is passed to a softmax classifier to estimate fault probabilities:

Gear defect detection starts by dividing the raw vibration signal s(t)$s(t)$ into smaller segments, each of which is converted into a spectrogram using the STFT for capture important frequency patterns. A CNN extracts spatial features from each spectrogram, while an LSTM processes these features over time to learn how vibration patterns evolve as gear faults develop. The final LSTM output is passed through a softmax classifier to predict the type of gear defect, and model trained by minimizing cross entropy loss between predicted and actual labels.

Result and Discussion

Architecture illustrated in **Fig. 3** represents hybrid **CNN LSTM model** designed for gear defect detection tasks. It begins with the **input layer**, where each vibration or acoustic signal segment in a sequence (denoted as $x_1, x_2... x_n$) is individually passed into the model. These segments may represent different time windows of the machine's operation. Each segment is first converted into a time-frequency representation (such as a spectrogram), which is then processed through a **Convolutional Neural Network (CNN) block**. This block extracts high level **spatial features** from spectrogram, like frequency patterns, harmonics,anomaly signatures indicative of specific gear defects (e.g., wear, cracks, or misalignment).The extracted feature maps are then fed into **LSTM block**, which processes features sequentially and captures **temporal dependencies** across the signal segments. This is crucial in gear defect detection, as faults often manifest through evolving patterns over time. The **output layer** utilizes the learned spatial-temporal representations to classify each input sequence or signal segment into predefined defect categories ($y_1, y_2, ..., y_n$). This hybrid CNN-LSTM architecture effectively combines spatial feature learning capabilities of CNNs with sequence modeling strength of LSTMs, resulting in an accurate and robust framework for real time gear condition monitoring and fault diagnosis.

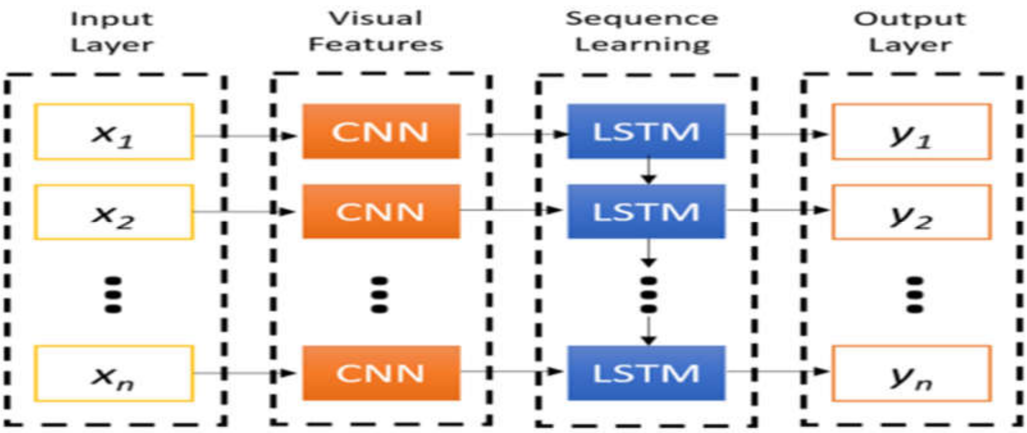

Fig 3 CNN-LSTM Architecture

Confusion matrix is table used evaluate performance of classification model. It shows how well model's predictions match the actual labels.

1. Precision

Precision is how many of predicted positive cases were actually correct.

Formula:

Precision = TP / (TP + FP)

High precision means fewer false positives.

2. Recall

Recall tells us how many of actual positive cases were correctly predicted.

Formula:

Recall = TP / (TP + FN)

High recall means fewer false negatives.

3. F1 Score

F1 Score is harmonic mean of precision and recall — it balances two.

Formula:

F1 Score = 2 * (Precision * Recall) / (Precision + Recall)

Especially useful when need a balance between precision and recall, and class distribution is uneven.

4. Support

Support is number actual occurrences of class dataset.

Formula:

Support = TP + FN

It is not performance metric but gives context the precision, recall, F1 score.

CNN Model Training VS Validation Accuracy

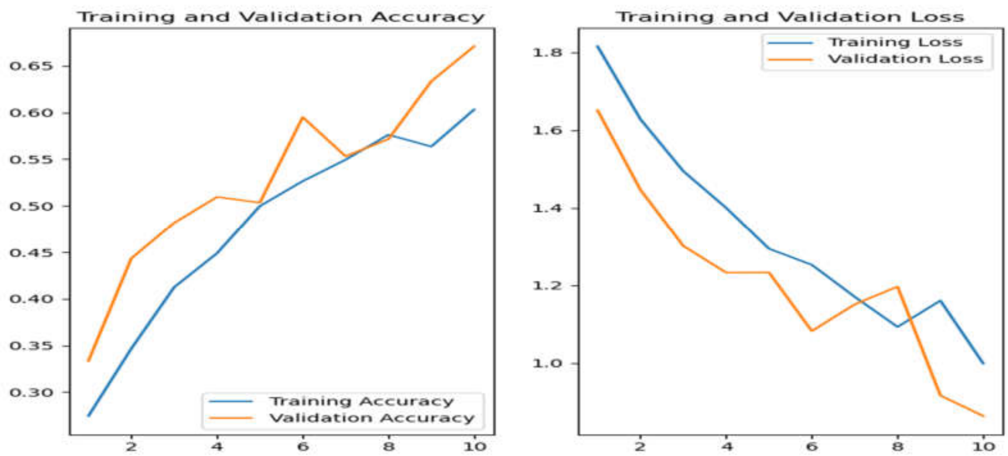

Fig 4 CNN Training VS Validation Accuracy, Loss

LSTM Model Training VS Validation Accuracy

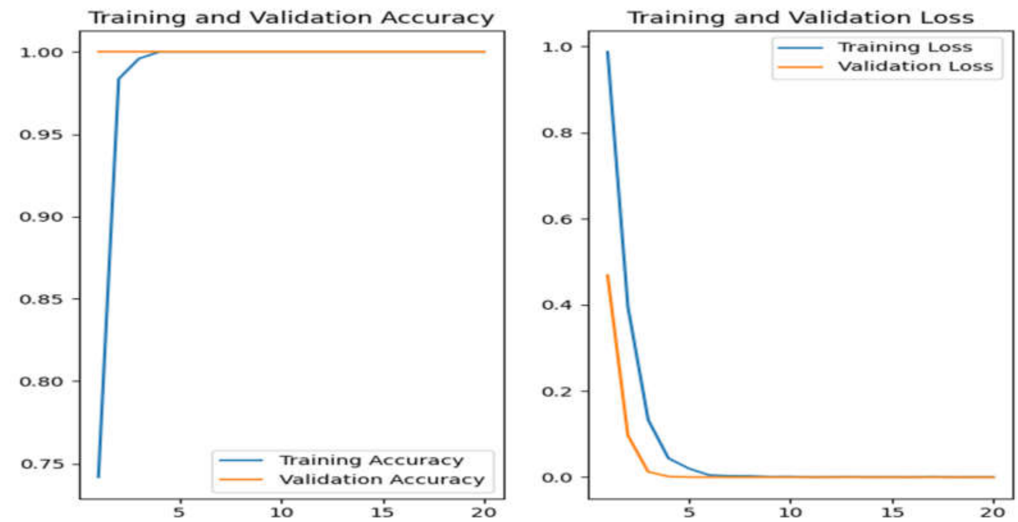

Fig 5. LSTM Training Accuracy Loss

CNN- LSTM Hybrid Model Accuracy VS Loss

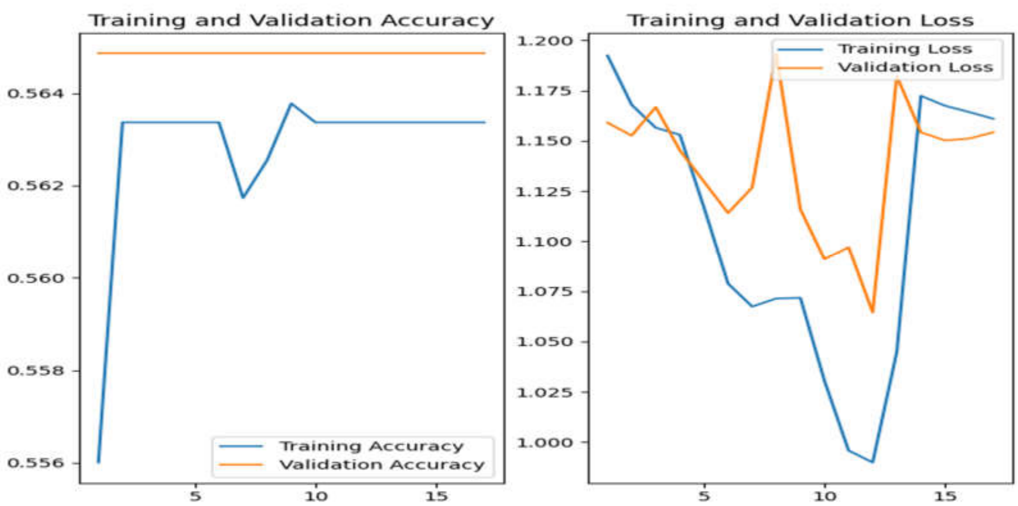

Fig 6 CNN-LSTM Training Accuracy, Loss

Gear Defect Detection

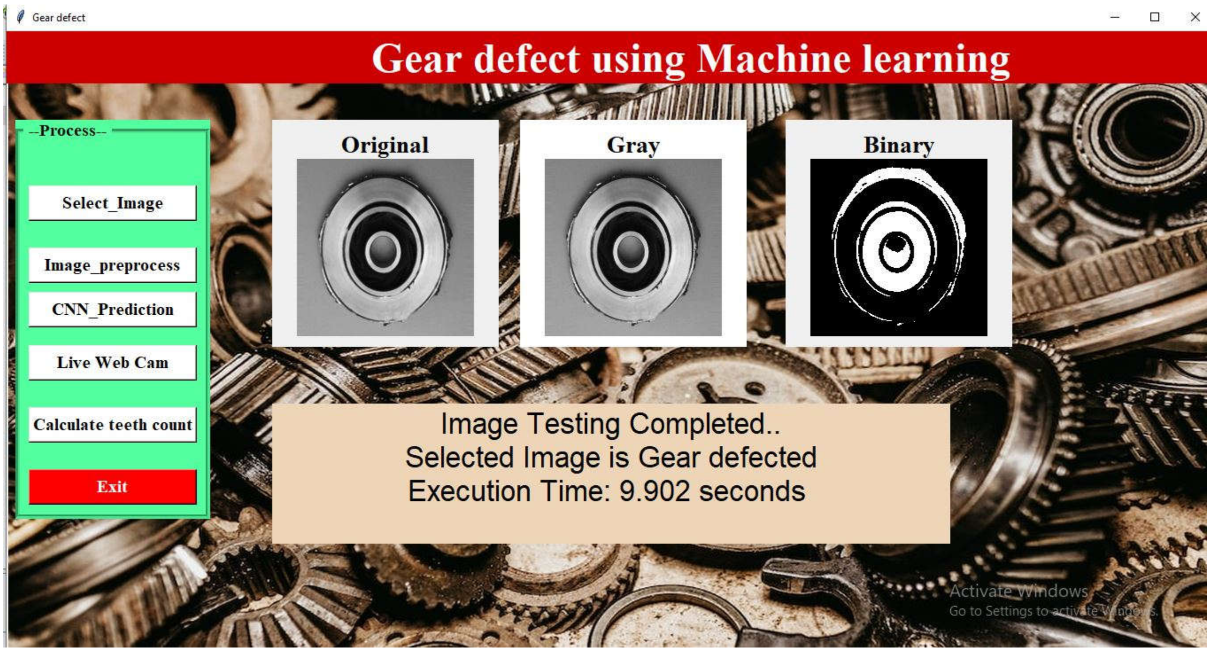

Fig 7 Prediction of Disease

Table 3 Performance analysis of CNN, LSTM, CNN-LSTM

Algorithm	Training Accuracy
CNN	0.80
LSTM	0.90
CNN-LSTM	0.92

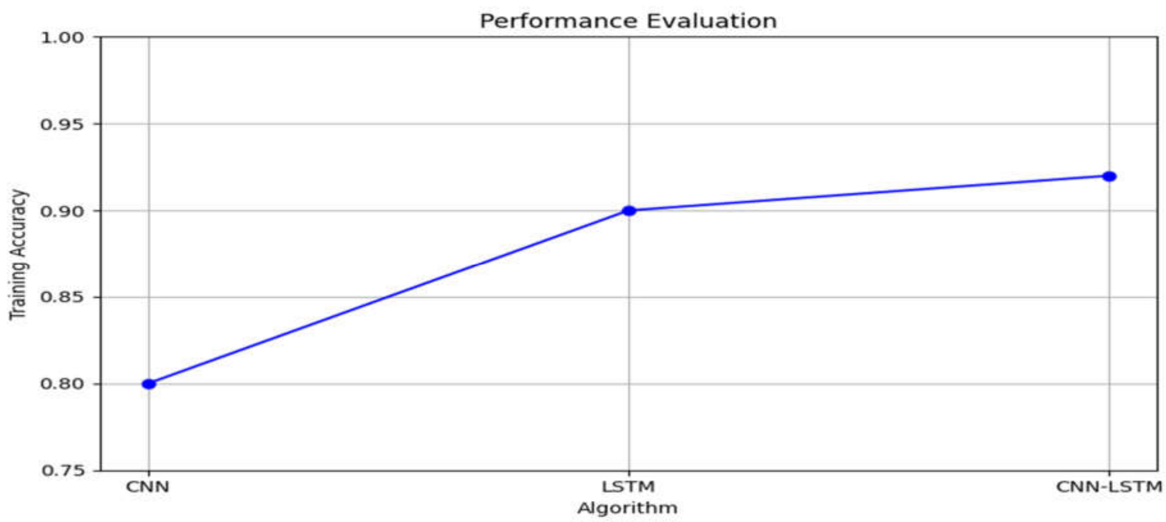

Fig 8 Performance Evaluation

Table 4 Evaluation metrics

Algorithm	Precision	Recall	F1 Score	Support
CNN	0.78	0.76	0.77	150
LSTM	0.88	0.89	0.88	150
CNN-LSTM	0.91	0.92	0.91	150

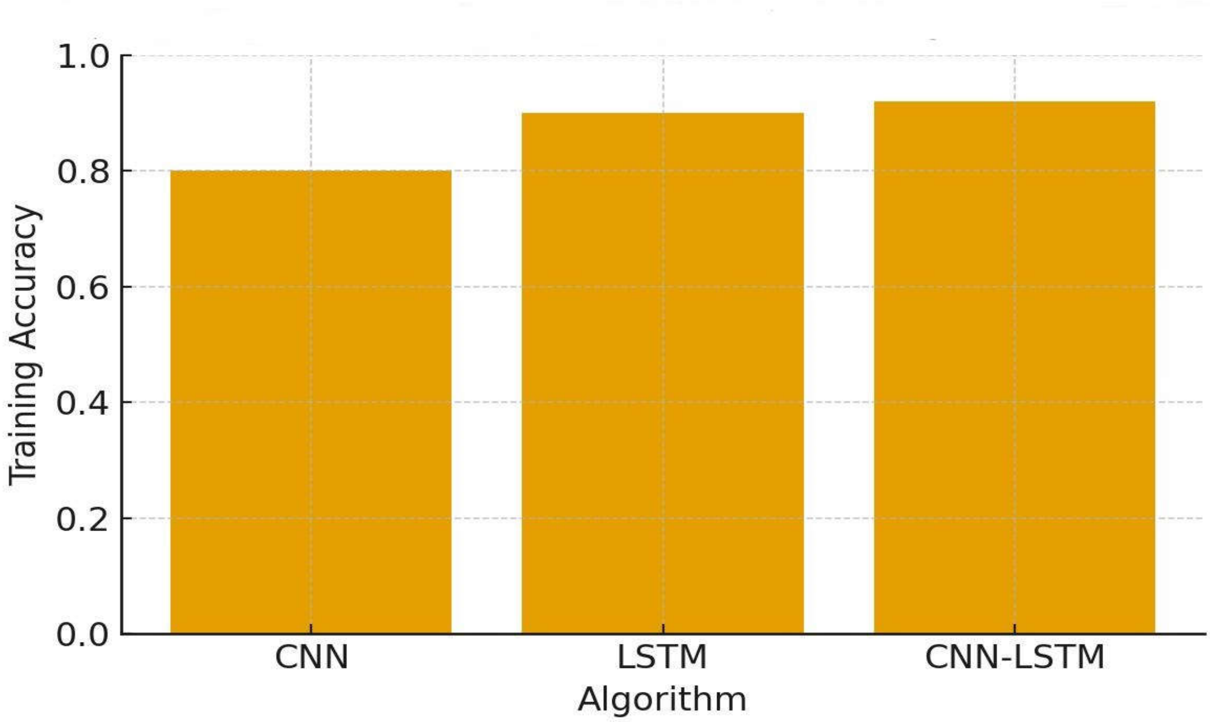

Fig. 9 Bar Chart of comparsion

Conclusion

This study presents hybrid DL framework for gear defect detection, integrating CN) with LSTM networks. CNN component excels at extracting spatial features from time frequency representations of vibration or acoustic signals, while LSTM component captures temporal dependencies, which are critical in understanding the progression and evolution of gear faults. The hybrid model effectively classifies various gear defects such as tooth wear, cracks, pitting, and misalignment with high accuracy and consistency. Confusion matrix analysis confirms model's capability differentiate between fault types with minimal misclassification. This intelligent diagnostic system offers significant benefits for predictive maintenance by enabling early fault detection and improving operational reliability. Looking ahead, model can be integrated into real-time monitoring systems on industrial machines using edge devices or IoT platforms. Further enhancements may include training on broader datasets, incorporating multisensor fusion, or adapting to other rotating machinery faults for broader applicability in industrial health monitoring.

Reference

1. Zhang, Y., Xu, X., & Zhang, S. (2020). A novel deep learning approach for gear fault diagnosis using CNN and LSTM. Mechanical Systems and Signal Processing, 138, 106550.

2. Gupta, R., Verma, A., & Soni, A. (2018). A comprehensive review of gear fault diagnosis using signal processing techniques. Mechanical Systems and Signal Processing, 107, 203-224.

3. Feng, Z., Tang, H.,& Liu, Z. (2019). Hybrid deep learning model for intelligent fault diagnosis of gearboxes. Journal of Sound and Vibration, 453, 58-74.

4. Hochreiter, S.,& Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735-1780.

5. LeCun, Y., Bengio, Y., \& Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.

6. Zhang, S., Xu, X., \& Li, Z. (2021). Deep learning-based gear fault detection using vibration signals and acoustic emission. Journal of Mechanical Engineering Science, 235(1), 120-130.

7. Application of deep CNN-LSTM network to gear fault diagnostics. Proceedings of the IEEE International Conference on Artificial Intelligence and Computer Applications, 9486591. https://doi.org/10.1109/ICAICA52286.2021.9486591.

8. An improved Conv-LSTM method for gear fault detection. In Advances in Artificial Intelligence and Security (pp. 110–121). Springer. https://doi.org/10.1007/978-3-031-20099-1_10

9. A distinguished deep learning method for gear fault diagnosis based on improved CNN-LSTM network. SN Applied Sciences, 6, Article 33. https://doi.org/10.1007/s42452-024-06033-7

10. Deep learning model for metal gear defect detection. In Proceedings of the International Conference on Artificial Intelligence and Computer Applications (pp. 189–200). Springer. https://doi.org/10.1007/978-981-96-2625-0_17

11. Deep learning used to detect gear inspection. Proceedings of the IEEE International Conference on Artificial Intelligence and Computer Applications, 10051817. https://doi.org/10.1109/ICAICA52286.2023.10051817

12. Helical gear defect detection system based on symmetrized dot pattern and convolutional neural network. Proceedings of the IEEE International Conference on Artificial Intelligence and Computer Applications, 10755078.

13. A hybrid approach for gearbox fault diagnosis based on deep learning and vibration signal analysis. The International Journal of Advanced Manufacturing Technology, 138, 73–85. https://doi.org/10.1007/s00170-024-13873-5.

14. Gear Assembly Fault Analysis Using Dual-Tree Complex Wavelet Transform Kashwan, K.R.,Senior Member. Deepika S [2016]

15. A Bevel Gear Quality Inspection System Based on Multi-Camera Vision Technology.Dexing Zhong2 and Jiuqiang Han4, Ruiling Liu1, Hongqiang Lyu3[2012]

16. Experiment research on the Metal Magnetic Memory in gear micro-crack detection. Lingli Cui2, Jianyu Zhang3, Lixin Gao4, and Yonggang Xu5,Chenhui Kang1[2015]

17. A Defect Status Detecting Method for External Gear in Railway.Fuqiang Li2,Yanfeng Li3,Xuyang Cao5,Chuangang Wang1, Houjin Chen4[2014]

18. Samarasinghe, T., Nguyen, A., \& Watson, S. (2018). Support vector machine for gear fault classification using time–frequency domain features. Mechanical Systems and Signal Processing, 110, 569–583. https://doi.org/10.1016/j.ymssp.2018.03.001

19. Zhao, R., Yan, R., Wang, J., Mao, K., \& Shen, F. (2017). Deep learning and its applications to machine health monitoring. Mechanical Systems and Signal Processing, 115, 213–237. https://doi.org/10.1016/j.ymssp.2018.05.050

20. Wang, H., Chen, P., Zhang, J., \& Wang, H. (2020). Gear fault diagnosis based on a hybrid CNN-LSTM model. IEEE Access, 8, 137316–137325. https://doi.org/10.1109/ACCESS.2020.3011234

21. Park, S., Kim, N., \& Lee, K. (2016). Fault diagnosis of gear systems using random forest algorithm. International Journal of Precision Engineering and Manufacturing, 17(8), 1081–1089. https://doi.org/10.1007/s12541-016-0134.