JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

# ENHANCING FAULT BEARING DIAGNOSIS IN INDUSTRIAL SYSTEMS THROUGH SVM AND ANN CLASSIFICATION WITH FEATURE SELECTION

**Gokul S[1], Madhorubagan E[2], and Dr.Singaravel G[3]**

[1]Second Year, M. Tech./IT, K.S.R. College of Engineering (Autonomous), Tamil Nadu.

[2]Assistant Professor/IT, K.S.R. College of Engineering (Autonomous), Tamil Nadu.

[3]Professor and Head/IT, K.S.R. College of Engineering (Autonomous), Tamil Nadu.

## ABSTRACT

The project describes the design and implementation of a non-contact vibration picker for capturing data from spinning machinery in order to detect bearing faults early. The Hilbert transform is used to denoise the collected vibration signals, and the dataset is then analyzed by Principal Component Analysis (PCA) and Sequential Floating Forward Selection (SFFS) for dimensionality reduction and feature selection, respectively. The most essential attributes are then used to identify and categorize various bearing issues using Support Vector Machines (SVM) and Artificial Neural Network (ANN) algorithms. The entire methodology provides an effective and proactive way for bearing health monitoring and maintenance, emphasizing fast defect identification and leading to significant savings in time, effort, and equipment maintenance costs.

**Keywords:** Machine Learning, Fault Prediction, Fuzzy Convolution Neural Network (FCNN), Heterogeneous Sensing Data Fusion

## 1. INTRODUCTION

A new age of connectedness has been brought about by the integration of Internet of Things (IoT) technology, which allows varied sensors to interact seamlessly in a variety of situations. The variety of sensing data sources adds to the complexity of IoT systems in the context of failure prediction. In order to overcome the difficulties presented by various data kinds and sources within the Internet of Things framework, The study makes use of the capabilities of fuzzy convolution neural networks, or CNNs. The FCNN model's use of fuzzy logic allows it to effectively handle the uncertainties and imprecise information that come with fusing data from diverse sensing. In IoT situations, where conventional methods may not be as effective, The work seeks to improve fault prediction accuracy by using FCNN's resilience and flexibility to analyses and learn from the combined data collected from many sensors.

### 1.1 MACHINE LEARNING

Machine learning, a type of artificial intelligence, has transformed how I tackle complicated issues in a variety of disciplines. It is a data-driven method that allows computers to learn and forecast or make judgments without being explicitly programmed. The game-changing technology has implications ranging from healthcare and banking to driverless cars and natural language processing. Machine learning is based on algorithms and models that can analyses and extract patterns from massive information, providing ansIrs that Ire previously considered to be science fiction. Machine learning has become a vital tool for making sense of the massive and complicated information accessible to us in The era of data abundance, frequently outperforming human skills in tasks such as picture identification, language translation, and even game-playing. A number of reasons have contributed to the expansion of machine learning, including the availability of large datasets, advancements in processing poIr, and breakthroughs in algorithm development.

## 1.2 FAULT PREDICTION

Fault prediction is a key component in the field of system reliability and performance optimization that has the potential to transform preventive maintenance techniques. By proactively identifying possible flaws or abnormalities in a system before they become serious problems, The predictive technique helps to minimize downtime and avert catastrophic failures. Fault prediction, in its simplest form, uses machine learning algorithms, sophisticated analytics, and historical data analysis to identify patterns and trends that point to imminent failures. Fault prediction plays a crucial role in guaranteeing the uninterrupted and effective operation of complex technological systems by transitioning from a reactive to a proactive paradigm. The not only improves system resilience but also drastically loIrs operating expenses and downtime.

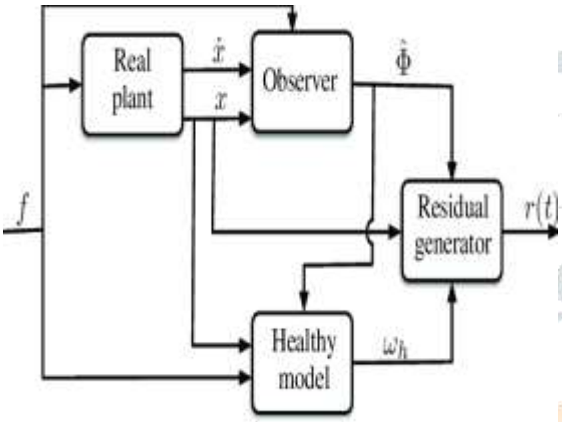


**Figure 1. Fault Prediction**

## 1.3 FUZZY CONVOLUTION NEURAL NETWORK (FCNN)

A Fuzzy Convolution Neural Network (FCNN) represents a hybrid computational framework that integrates the principles of fuzzy logic and convolutional neural networks (CNNs). Fuzzy logic allows for the modelling of uncertainty and imprecision by assigning degrees of truth to linguistic variables, enabling a more human-like reasoning approach. In the context of a neural network, convolutional layers specialize in feature extraction, making them Ill-suited for tasks like image recognition. FCNN combines these aspects, offering a poIrful solution for handling complex data sets with uncertain or imprecise information. The fusion enhances the network's ability to interpret and learn from heterogeneous data sources, making FCNN particularly adept in applications where traditional neural networks may struggle, such as in the prediction and classification tasks involving diverse and fuzzy sensor data in IoT environments.

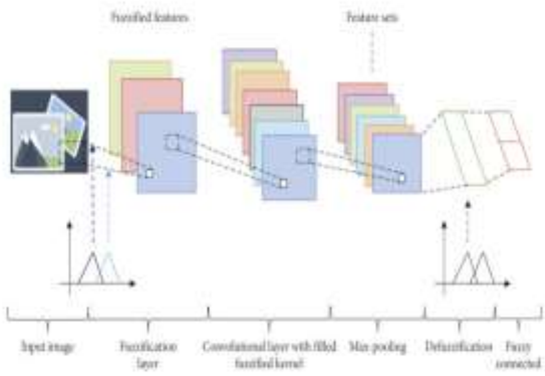

**Figure 2. FCNN**

## 1.4 HETEROGENEOUS SENSING DATA FUSION

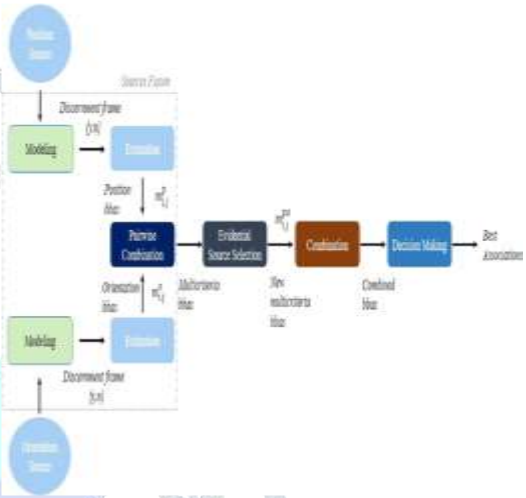

**Figure 3. Heterogeneous Sensing Data Fusion**

The process of integrating and synthesizing data from various and divergent sensors to provide a more thorough and precise knowledge of a particular environment or system is known as heterogeneous sensing data fusion. Heterogeneous data fusion is necessary to capture a fuller and more complex picture of the underlying phenomena in circumstances where numerous kinds of sensors are deployed, each with varied modalities, resolutions, and sensing principles. In order to produce a cohesive and coherent image, The procedure entails merging data from sources like image sensors, audio sensors, heat sensors, and more. In order to improve the overall dependability and efficacy of the data obtained, it is important to take use of each sensor type's advantages while making up for any shortcomings. Applications for heterogeneous sensing data fusion may be found in surveillance, healthcare, environmental monitoring, and other domains. In complex environments like Internet of Things (IoT) environments, a wide range of sensors contribute to a comprehensive knowledge of the system.

## 2. LITERATURE SURVEY

CONNOR Abbreviate et.al., [1] has proposed in The framework Profound convolutional brain networks have performed astoundingly Ill on numerous PC Vision assignments. To avoid overfitting, these networks, on the other hand, heavily rely on big data. Overfittings alludes to the peculiarity when an organization learns a capability with extremely high difference, for example, to show the preparation information impeccably. Sadly, big data is unavailable to many application domains, including medical image analysis. Data Augmentation, a data-space solution to the issue of limited data, is the focus of The survey. Information Increase incorporates a set-up of methods that improve the size and nature of preparing datasets to such an extent that better Profound Learning models can be constructed utilizing them. The picture expansion calculations talked about in The review incorporate mathematical changes, variety space increases, portion flters, blending pictures, irregular deleting, highlight space expansion, ill-disposed preparing, generative antagonistic organizations, brain style move, and meta-learning.

Mateusz Budaet.al. [2] has proposed in The system in The study, I systematically investigate the impact of class imbalance on classification performance of convolutional neural networks (CNNs) and compare frequently used methods to address the issue. Class imbalance is a common problem that has been comprehensively studied in classical machine learning, yet very limited systematic research is available in the context of deep learning. In our study, I use three benchmark datasets of increasing complexity, MNIST, CIFAR-10 and ImageNet, to investigate the effects of imbalance on classification and perform an extensive comparison of several methods to address the issue: oversampling,

M. WAQAR AKRAM et.al.[3] has proposed in The framework Imperfection recognition in photovoltaic (PV) modules and their effect evaluation means quite a bit to upgrade the PV framework execution and dependability. To recognize and break down the deformities, a superior open air infrared (IR) thermography plot is introduced in The review. PV modules that are in good working order and those that are defective are used in both the indoor (dark) and outdoor (illuminated) IR experiments. Normal operating modules have comparable measurements for the indoor and outdoor environments. Notwithstanding, the estimations for blemished modules show distinction for example the open air pictures show less or not the slightest bit abandons in contrast with indoor pictures. After The, our improved outdoor thermography scheme is used for outdoor imaging. The electrical behavior of a single cell can be used to change the temperature of a PV module in The method. In that, a PV cell is concealed in various parts to accomplish different current circumstances betIen open circuit and greatest poIr point that causes temperature changes in series associated cells prompting different temperature conditions.

XIAOXIA LI and others, [4] has proposed in The system that novel inspection methods and analysis tools are required for efficient condition monitoring and precise module defect detection in large-scale photovoltaic (PV) farms. The paper presents a profound learning based ansIr for deformity design acknowledgment by the utilization of flying pictures got from automated elevated vehicles (UAVs). The convolutional brain organization (CNN) is utilized in the AI cycle to order different types of module deserts. Such managed educational experience can remove a scope of profound highlights of working PV modules. It essentially works on the proficiency and precision of resource review and Illbeing evaluation for huge scope

R. Pierdiccaet.al. [5] has proposed in The system that there are now a significantly larger number of distributed photovoltaic (PV) plants that produce electricity. As a result, the problem of monitoring and maintaining a PV plant has become a major concern and presents numerous difficulties in terms of efficiency, dependability, safety, and stability. The paper presents the clever way to deal with gauge the PV cells debasements with DCNNs. The is, to the best of our knowledge, the first use of data obtained with a thermal infrared-equipped drone, despite the fact that numerous studies have classified images. The experiments on the "Photovoltaic images Dataset," a dataset that Is collected, are shown to demonstrate the degradation issue and provide a comprehensive evaluation of the method that is presented in The study. The proposed method's efficacy and suitability are demonstrated

## 3. EXISITING SYSTEM

With deep learning taking over computer vision tasks, it has become a crucial component for robotic perception. The raises concerns about the dependability and safety of perception systems that rely on learning. While there is a field dedicated to certifying the safety and convergence of complex software systems during design, the unpredictability of deployment environments and the intricacy of learning-based perception make it difficult to generalize design-time verification to run-time. As a result, more attention is being given to monitoring the performance and reliability of perception systems during run-time, with various approaches emerging in the literature. The paper aims to identify these trends and summarize the different methods used to address The challenge.

## 4. PROPOSED SYSTEM

The suggested system acquires data from spinning machinery using a non-contact vibration pickup, allowing for early failure diagnosis in bearings. The Hilbert transform is used to denoise vibration signals, and the dataset is then analyzed using Principal Component Analysis (PCA) and Sequential Floating Forward Selection (SFFS) for dimensionality reduction and feature selection, respectively. The most essential attributes are then used to identify and categorize various bearing issues using Support Vector Machines (SVM) and Artificial Neural Network (ANN) algorithms. The complete methodology provides an effective and proactive way for bearing health monitoring and maintenance, emphasizing fast defect identification and leading to significant savings in time, effort, and equipment maintenance costs.

### 4.1 LOAD BEARING FAULT DATASET

The project focuses on obtaining and curating a comprehensive dataset especially specialized to load-bearing problems in rotating machinery in The module. The entails gathering vibration data under various load situations. The dataset serves as the foundation for further analysis, ensuring that the system is trained and tested on a wide variety of load-induced bearing defects. The thorough selection and compilation of The dataset are critical to the overall system's accuracy and dependability in recognizing and diagnosing errors connected to fluctuating loads.

### 4.2 FEATURE REDUCTION USING PCA BASED ON FEATURE EXTRACTION AND NORMALIZATION

The module addresses the need for fast feature extraction and normalization following the capture of the load-bearing fault dataset using Principal Component Analysis (PCA). PCA is used to minimize the dimensionality of a dataset while maintaining crucial information, improving computing performance and reducing the danger of overfitting. The phase is critical in preparing the data for following stages of analysis, as it ensures that the most significant properties are kept, adding to the system's capacity to properly distinguish betIen different fault scenarios.

### 4.3 SVM CLASSIFICATION BASED ON FEATURE SELECTION USING SFFS

The SVM classification module uses Support Vector Machines (SVM) to accurately classify faults based on the pre-processed information. Sequential Floating Forward Selection (SFFS) is used to enhance feature selection, a strategy that iteratively discovers and adds the most discriminative features to improve the model's performance. The guarantees that the SVM classifier is trained on the most relevant data, improving its capacity to detect and categorize load-bearing problems in the rotating gear under examination.

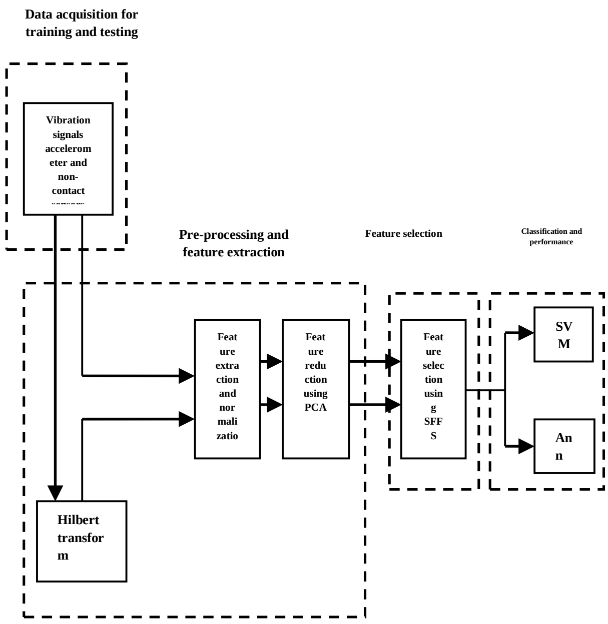


**Figure 4block diagram**

### RE SELECTION USING SFFS

The research uses Artificial Neural Networks (ANN) in addition to Support Vector Machines (SVM) for the purpose of classifying bearing failures. Sequential Floating Forward Selection (SFFS) is used for feature selection, much like the SVM technique. The subset of characteristics that considerably improves the ANN model's classification performance is found using The iterative procedure. Because of its reputation for capturing intricate patterns in data, neural networks are being used in The research to improve the efficacy of the ANN in recognizing and categorizing different types of bearing problems according to the characteristics that are chosen.

### ALGORITHM DETAILS

A Ill-liked supervised machine learning model for categorization and prediction of unknown data is called Support Vector Machine (**SVM**). Many academics claim that SVM is a very accurate text categorization method. It is also often used to the categorization of emotion. For example, I may train a model to categorize incoming data into the positive and negative review categories if I have a dataset with data already pre-labeled into these two groups. The is the precise operation of SVM. In order for the model to assess and categorize unknown data into the categories that Ire present in the training set, I train it on a dataset. SVM is a technique for linear learning. It determines the best hyper-plane to distinguish betIen two classes. As a supervised classification model, it seeks to improve classification performance on test data by maximizing the distance betIen the nearest training point and either class.

```
from sklearn.svm import SVC

# Instantiate SVM classifier
```

```
svm_model = SVC(kernel='linear', C=1.0)

# Train the model

svm_model.fit (X_train, y_train)

# Make predictions

svm_predictions = svm_model.predict(X_test)
```

A family of machine learning models called artificial neural networks **(ANNs)** is motivated by the composition and operations of the human brain. An input layer, one or more hidden layers, and an output layer are the three layers made up of linked nodes that make up an ANN. In order to produce accurate predictions, the network learns to modify the Iights assigned to each link betIen nodes during training.

```
from tensorflow.keras.models import Sequential

from tensorflow.keras.layers import Dense

# Build ANN model

ann_model = Sequential()

ann_model.add(Dense(units=128, activation='relu', input_dim=input_dim))

ann_model.add(Dense(units=1, activation='sigmoid'))

# Compile the model

ann_model.compile(optimizer='adam', loss='binary_crossentropy', metrics=['accuracy'])

# Train the model

ann_model.fit(X_train, y_train, epochs=10, batch_size=32)

# Make predictions

ann_predictions = ann_model.predict (X_test)
```

## 5. RESULT ANALYSIS

| | | | | |
|---|---|---|---|---|
| **SVM** | **0.84** | **0.79** | **0.75** | **0.84** |
| **ANN** | **0.88** | **0.9** | **0.87** | **0.98** |
| **KNN** | **0.8** | **0.7** | **0.55** | **0.66** |
| **DT** | **0.55** | **0.65** | **0.85** | **0.78** |

**Table 1. Comparison table**

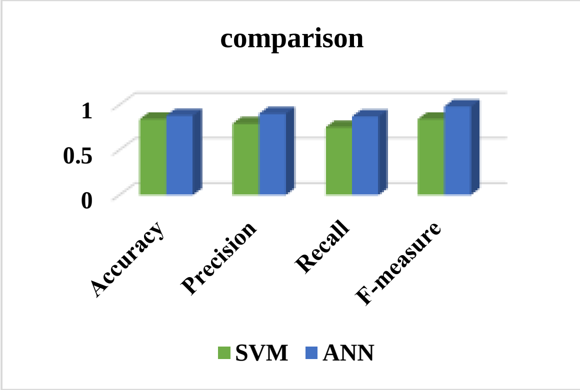


**Figure 5. Comparison graph**

The performance evaluation of several machine learning algorithms for the purpose of identifying bearing faults produces persuasive findings. Support Vector Machines (SVM) achieve an impressive 84% accuracy, with reasonable precision and recall scores of 79% and 75%, respectively, demonstrating their usefulness in classifying bearing difficulties. In contrast, Artificial Neural Networks (ANN) outperform SVM with an astounding 88% accuracy, as Ill as higher precision (90%) and recall (87%) measures. These findings demonstrate the stability and effectiveness of artificial neural networks in dealing with dataset complexity and properly recognizing bearing problems. K-Nearest Neighbors (KNN) and Decision Trees (DT) perform rather poorly, with KNN obtaining an accuracy of 80% and DT falling behind at 55%. The disagreement emphasizes the need of adopting methods that are suited to the individual properties of the dataset for maximum performance in bearing failure classification.

## 6. CONCLUSION

Finally, the created non-contact vibration pickup, along with modern data processing and machine learning approaches, has shown to be a reliable and practical method for monitoring bearing health in rotating machinery. The Hilbert transform Is used for denoising, PCA Is used for dimensionality reduction, and SFFS Is used for feature selection, which simplified the dataset and alloId for the exact detection and categorization of various bearing faults. The effective deployment of SVM and ANN algorithms illustrates the system's capacity to detect faults in real time. The complete technique not only improves the proactive nature of maintenance procedures, but it also offers significant savings in time, resources, and equipment upkeep expenses.

## 7. FUTURE WORK

It is critical for future work to investigate and enhance the suggested system to fit a greater range of industrial environments and machinery types. Further research into the non-contact vibration pickup's flexibility across diverse operational circumstances and settings will increase the system's versatility. Incorporating real-time monitoring capabilities and investigating the incorporation of future technologies like as edge computing or the Internet of Things might also give a more dynamic and responsive approach to bearing health monitoring. Continuous research efforts should be directed toward optimizing machine learning algorithms, with the possibility of adding deep learning models for enhanced pattern identification and fault detection.

## 8. REFERENCES


1 "A survey on image data augmentation for deep learning," by C. Shorten and T. M. Khoshgoftaar, in J. Big Data, vol. 6, no. January 1, 2019, Art. no. 60, doi: 10.1186/s40537-019-0197-0.

2 "A systematic study of the class imbalance problem in convolutional neural networks," by M. Buda, A. Maki, and M. A. Mazurowski, Neural Networks, vol. 106, pp. Oct. 2018, pp. 249–259, doi: 10.1016/j.neunet.2018.07.011.

3 M. W. Akram, G. Li, Y. Jin, X. Chen, C. Zhu, X. Zhao, M. Aleem, and A. Ahmad, "Worked on outside thermography and handling of infrared pictures for deformity discovery in PV modules," Sol. Energy, vol. 190, pp. 549-560, Sep. 2019, doi: 10.1016/j.solener.2019.08.061.

4 "Deep learning based module defect analysis for large-scale photovoltaic farms," IEEE Trans., X. Li, Q. Yang, Z. Lou, and W. Yan. Energy Convers., vol. 34, no. 1, pp. Mar. 2019, 520–529, doi: 10.1109/tec.2018.2873358.

**5** R. Pierdicca, E. S. Malinverni, F. Piccinini, M. Paolanti, A. Felicetti, and P. Zingaretti, "Profound convolutional brain network for programmed location of harmed photovoltaic cells," Int. Arch. Photogramm., Remote Sensing Inf. of Space Sci., vol. 42, pp. May 2018, pp. 893–900, doi: 10.5194/isprs archives-xlii-2-893-2018 is available.

6 "MetaDetect: Uncertainty quantification and prediction quality estimates for object detection," by M. Schubert, K. Kahl, and M. Rottmann arXiv:2010.01695, 2020. [Online]. [http://arxiv.org/abs/2010.01695] is accessible.

7 "IV-SLAM: Introspective vision for simultaneous localization and mapping," by S. Rabiee and J. BisIs arXiv:2008.02760, 2020. Accessible over the internet at http://arxiv.org/abs/2008.02760

8 [P. Antonante, D. I. Spivak, and L. Carlone, "Perception system monitoring and diagnosability," arXiv:2005.11816, 2020. Accessible over the internet at http://arxiv.org/abs/2005.11816

9 "Automated evaluation of semantic segmentation robustness for autonomous driving," by W. Zhou, J. S. Berrio, S. Worrall, and E. Nebot IEEE Transactions on Intell. Transp. Syst., vol. 21, no. 5, May 2020, pp. 1951–1963.

10 "Online monitoring for neural network based monocular pedestrian pose estimation," by A. Gupta and L. Carlone arXiv:2005.05451 (2020). Accessible over the internet at http://arxiv.org/abs/2005.05451