Results in Control and Optimization 17 (2024) 100491

Contents lists available at ScienceDirect

Results in Control and Optimization

journal homepage: www.elsevier.com/locate/rico

Noise robust classification of carbide tool wear in machining mild steel using texture extraction based transfer learning approach for predictive maintenance

Ravi Sekhar [a], Sharnil Pandya [b], Pritesh Shah [a,*], Hemant Ghayvat [b], Deepak Sharma [c], Matthias Renz [c], Deep Shah [d], Adeeth Jagdale [e], Devansh Hukmani [f], Santosh Saxena [a], Neeraj Kumar [g,h,i,j]

[a] Symbiosis Institute of Technology, Pune Campus, Symbiosis International (Deemed University) (SIU), Pune 412115, Maharashtra, India
[b] Department of Computer Science and Media Technology, Linnaeus University, Vaxjo, Sweden
[c] Department of Computer Science, Christian-Albrechts-University Kiel, Kiel, 24118, Germany
[d] University of Exeter, Exeter, United Kingdom
[e] Shree Ganesh Enterrprises, MIDC Chinchwad, Pune 411019, Maharashtra, India
[f] College of Professional Studies- Analytics, Northeastern University, 360 Huntington Avenue, Boston, MA 02115, USA
[g] Neeraj Kumar is CSED, Thapar Institute of Engineering and Technology, Patiala, Punjab, India
[h] School of Computer Science, University of Petroleum and Energy Studies, Dehradun, Uttarakhand, India
[i] Department of Electrical and Computer Engineering, Lebanese American University, Beirut, Lebanon
[j] Faculty of Computing and IT, King Abdulaziz University, Jeddah, Saudi Arabia

ARTICLE INFO

Keywords:
Smart manufacturing
Acoustics
Sound classification
Predictive maintenance
Deep learning
Transfer learning
Texture extraction
Tool condition monitoring

ABSTRACT

Acoustics based smart condition monitoring is a viable alternative to mechanical vibrations or image-capture based predictive maintenance methods. In this study, a texture analysis based transfer learning methodology was applied to classify tool wear based on the noise generated during mild steel machining. The machining acoustics were converted to spectrogram images and transfer learning was applied for their classification into high/medium/low tool wear using four pre-trained deep learning models (SqueezeNet, ResNet50, InceptionV3, GoogLeNet). Moreover, three optimizers (RMSPROP, ADAM, SGDM) were applied to each of the deep learning models to enhance classification accuracies. Primary results indicate that the InceptionV3-RMSPROP obtained the highest testing accuracy of 87.50%, followed by the SqueezeNet-RMSPROP and ResNet50-SGDM at 75.00% and 62.50% respectively. However, SqueezeNet-RMSPROP was determined to be more desirable from a practical machining quality and safety perspective, owing to its greater recall value for the highest tool wear class. The proposed acoustics-texture extraction-transfer learning approach is especially suitable for cost effective tool wear condition monitoring involving limited datasets.

1. Introduction

Industrial competitiveness of an enterprise depends upon the productivity of its machinery. Machine breakdowns and inoperability due to part wear, component failures and/or scheduled maintenance result in productivity losses. Smart manufacturing principles stipulate implementation of online machine tool condition monitoring and diagnostics. Smart manufacturing is a necessity

* Correspondence to: Symbiosis Institute of Technology, Pune Campus, Symbiosis International (Demmed University), Pune 412115, India.
E-mail address: pritesh.shah@sitpune.edu.in (P. Shah).

https://doi.org/10.1016/j.rico.2024.100491
Received 22 September 2023; Received in revised form 17 September 2024; Accepted 9 November 2024
Available online 19 November 2024
2666-7207/© 2024 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

for modern industries to survive and stay competitive. New technology deployment can help enterprises minimize losses and enhance quality output. A typical manufacturing cell emanates numerous acoustic events which generally tend to be ignored by operators/supervisors as they focus on their immediate job at hand. Proper detection, analysis and classification of such machining sounds with regards to machining conditions can lead to significant process improvements and time savings. Fellow researchers have conducted experiments in this research area; however, acoustic based event detection and classification towards predictive maintenance cutting tools in mild steel machining is an open research problem. Generally, vibration sensors are used for online machine performance data collection, however the tools required for the subsequent analysis of vibration data are often complex and costly [1]. Acoustic event based machining noise data collection and analysis tools overcome these limitations of the vibrations based online monitoring architecture [2].

In the context of acoustics based tool condition classification problems, machine learning and/or other similar methods are generally employed by researchers. Hence, there is a lot of scope to explore the application of various deep learning architectures in this field. Moreover, scenarios involving limited datasets necessitate advanced techniques such as transfer learning to be considered in the scope of investigations. In the present study, acoustic events were solely considered for tool condition monitoring, unlike many research studies wherein acoustics have been considered in conjunction with other sensor data (thermal, vibrations). Recording acoustic events is generally more cost effective than other sensors, and hence this methodology ensures wider applicability of non evasive tool condition monitoring. Deep learning architectures can capture more feature dynamics in an acoustic signature as compared to machine learning and other similar methods, but these algorithms typically require a lot of training data to yield satisfactory results. Herein, transfer learning enables deep learning algorithms to be applied to relatively smaller datasets with reasonably better accuracy, thus making acoustics based tool condition monitoring much more practically implementable. Recently, many researchers have investigated acoustic event analysis for predictive maintenance of industrial equipment, as detailed in the following subsection.

1.1. Noise robust classification of mechanical systems using machine learning

The past few years have witnessed significant research efforts directed towards acoustic signals based fault classification in many mechanical systems. Kane and Andhare [1] classified gearbox faults based on acoustic, psychoacoustic and vibration emissions during line inspection. Gear faults were mapped to vibration / acoustic signals using discriminant qualifier and ANN techniques. The pyschoacoustic feature was found to attain best accuracy of fault detection. However, the study did not include transfer learning approach, which is ideally suited for classification problems involving few data samples. Kene and Choudhury [3] proposed an analytical fusion model to combine force, surface roughness and vibration sensor data for better tool monitoring. However, acoustic feedback based tool condition monitoring was not explored, which is easier to implement as copared to vibration sensor based monitoring. Bhuiyan and Choudhury [4] reviewed sensors and sensing techniques in tool condition monitoring applications. Artificial intelligence and machine learning based tool condition monitoring was identified as one of the possible future research directions in this field. Alatorre et al. [5] implemented closed loop control of force system during robotic grinding of nickel alloy beams using sound signals. However, advanced machine learning/deep learning techniques could not be incorporated for tool condition classifications. Lee et al. [6] applied artificial intelligence for predictive maintenance of spindle motor and cutting tool in machining. However, the authors did not explore deep learning architectures for tool condition monitoring in their study. Nekane et al. [7] developed an autonomous ultrasonic predictive maintenance system for wind turbines. However, ultrasonics based condition monitoring may prove too costly for a regular machining tool wear classification. Khan et al. [8] utilized airborne sound emissions to model and predict gear shaft misalignment under dynamic loading conditions. This study did not focus on gear wear identification, however it is a good example of acoustics based equipment maintenance. Becht et al. [9] combined microphone and structural sensors to detect loose bolts in a complex beam assembly. They added artificial noise to sensor measurements to demonstrate satisfactory performance of the acoustic-vibration sensors in presence of high ambient chatter. This study also highlighted the versatile applicability of acoustics in monitoring integrity of structures. However, deep learning based classification was not explicitly carried out. Delvecchio et al. [2] reviewed latest methodologies in acoustic based condition monitoring of internal combustion engines. They confirmed that acoustic based fault detection is superior to vibration assisted monitoring due to their non evasive nature. The authors also furnished guidelines for selection of appropriate signal processing techniques based on the type of system malfunctions being detected. This study conclusively proved the applicability and suitability of acoustics over vibration based anomaly classification in mechanical systems. However, acoustics based anomaly detection is not always easier as shown by Schnabel et al. [10], who concluded that acoustic events can detect plastic deformation in roller bearings due to particle contamination only at low speeds. They found that transient force signals mask plastic deformation signals at high speeds, making their acoustic detection difficult. On the other hand, Martin et al. [11] applied unsupervised feature machine learning to detect and classify lubricant contaminants in high speed roller bearings. However, deep learning techniques were not explored by the authors. Fuentes et al. [12] applied Gaussian mixture models to detect incipient roller bearing damage in wind turbines before the damage developed further to affect gear box resulting in turbine downtime. This study demonstrated how acoustic signals could be used for preventive maintainence. However, the authors did not apply suitable deep learning algorithms for enhanced classification results. Dykas et al. [13] applied instantaneous RMS acoustic signals to detect valve closure, combustion and fuel injector malfunction in a single cylinder diesel engine setup. The study did not involve machine learning or deep learning algorithms for malfunction classifications. Dias et al. [14] analysed harmonic component of acoustic emissions in centreless grinding to characterize resultant cylindricity, roundness and surface roughness. This study relied on fast fourier transform and wavelet analysers for classification of acoustic signals. Machine and/or deep learning were not explored, which can prove simpler and cheaper techniques in this field,

if applied correctly. Becht et al. [15] implemented the Time Reversal Multiple Signal Classification algorithm to suggest optimum sensor placement to localize damage in an aluminium framework. However, this study did not explore acoustic-textures, which make wear classification much easier to implement.

Recently, Twardowski et al. [16] applied decision tree approach to monitor the tool edge condition of a milling cutter while machining aluminium-silicon carbide metal matrix composite material. The authors reported an error of less than 6% was achieved while predicting tool condition classification. This study is an excellent example of machine learning application to a machining tool wear classification. However, the study did not explore the more advanced deep learning algorithms for the same. Straka and Corny [17] identified the occurrence of geometric errors during wire electrical discharge machining (WEDM) of circular profiles using acoustic emissions. While this study correlated wire tool electrode vibrations to the corresponding acoustic emissions for better machined surface quality, it did not employ machine or deep learning for dimensional inaccuracy classification based on acoustics. Cooper et al. [18] applied generative adversarial neural network to detect anomalies in the time-frequency domain of acoustic spectra obtained during milling process. In a related work, the authors [19] applied sonic signals based convolutional neural networks to monitor tool condition during vertical milling. However, these studies did not include a comparative transfer learning approach involving different deep learning algorithms, which is suitable for achieving effective classifications in case of limited datasets. Wan et al. [20] applied random forest feature combination and optimization to classify different wear states of ceramic grinding wheels using acoustic emissions. The authors showed how frequency based feature bands enable better wear identification accuracies. However, acoustics-texture based transfer learning approach was not explored on different deep learning algorithms-optimizers combinations for comparative results. Xu et al. [21] used energy principle to establish relationship between robotic grinding belt wear and the corresponding acoustic signals for accurate belt wear prediction, with an average error of approximately 10 %. However, the study did not explore advanced machine/deep learning methodologies to achieve even better prediction accuracy. Revill et al. [22] used acoustic signals to propose a novel structural health monitoring system for self lubricating composite bearing liners. The study did not explore acoustic-texture based wear classification using various deep learning approaches for enhanced prediction accuracies. Shen [23] used support vector machine algorithm with linear kernel to classify surface grinding wheel wear using machining noise data. The authors extracted data features from the acoustic signals using time and frequency domain representations. This study also did not consider transfer learning based exploration of various deep learning-optimizer architectures for enhanced classification of tool wear. Deshpande et al. [24] applied several machine learning algorithms to classify acoustic emission features corresponding to different wear rates generated during tribological testing. This study compared acoustics based wear classification accuracies obtained by multiple machine learning algorithms, but did not include deep learning architectures for the same. Unterberg et al. [25] applied feature based analysis to monitor fine blanking tool wear using acoustic emissions. This study applied feature engineering to visualize acoustic signals corresponding to various tool wear modes. However, machine learning/deep learning based classification was not carried out. Kuntoglu and Saglam [26] used fuzzy rule system to predict the tool flank wear, acoustic emissions and cutting temperature based on machining parameters. Transfer learning was not considered as a special case to handle classification problems involving limited datasets. Konig et al. [27] used long short-term memory technique to monitor sliding bearing wear using acoustic emission signals for wind turbine applications. This study did consider recurrent neural networks for acoustic signal classification, but deep learning algorithms were not in scope. Ahmed et al. [28] applied wavelet transform to filter acoustic signals generated while machining AISI 304 stainless steel. Acoustic emissions were utilized to monitor, classify and predict built up edge formation on tool surfaces. This study did not consider acoustic signals exclusively, as it incorporated cutting force signals data as well for built up edge classification. A pure acoustics-based approach was deemed necessary to fully analyse its potential for tool wear classification in machining, for wider applicability.

1.2. Motivation, aim and scope of the present study

The above mentioned literature review suggests that most studies have focused on acoustics and machine learning based monitoring of gear boxes, wind turbines, gear shafts, bolts, diesel engines, milling, WEDM, grinding, belts, bearings and tribological testing applications. Fewer studies seem to have applied the same on tool wear classification during turning of mild steel rods, a widely used industrial structural material. Secondly, studies need to investigate various deep learning architectures to achieve accurate tool wear detection and classification through texture-based analysis of acoustic emissions. Thirdly, there is a need to investigate the application of transfer learning methodology for acoustic emission-texture based classification of tool wear, especially when the number of samples for training are very few. Hence, the present work aimed to bridge these gaps by employing acoustic emission-texture based transfer learning methodology to detect, classify and predict tool wear while machining mild steel using carbide inserts. This approach will enable manufacturing firms to monitor tool condition without stopping/pausing production, since the acoustic based condition monitoring will eliminate the need for physical tool insert checking for wear/damage. Moreover, acoustic based condition monitoring will enable manufacturing firms to take better informed predictive actions such as tool regrinding/replacement and prevent/minimize machine downtime/production stoppages. The following subsections briefly present a general overview of the development of machine learning methods into deep learning techniques, followed by a short introduction to the transfer learning methodology.

1.3. Classification using deep learning

Machine learning finds widespread applications in system modelling and prediction across various domains such as lean manufacturing [29,30], robotics [31], fault diagnosis [32], composite machining [33], machine condition monitoring [34], electric

vehicles [35], [36] topic modelling [37,38] and many more. In machine learning, the convolutional neural networks (ConvNets) have been widely applied in diverse applications, helped by the ever increasing computing power and the data available for training [39]. From the year 2014 onward, the existing ConvNets were evolved to develop deeper, many layered convolutional networks for enhanced classification applications [40]. Deep learning models such as VGGNet [41] and GoogleLeNet [42] brought revolutionary advances in classification applications including object detection [43], object tracking [44], segmentation [45] and superresolution [46], classification of videos [47] and human pose [48], among many others. The deep learning models were able classify data based on such finer features which were difficult/not possible for machine learning methods to successfully identify [49].

However, the deep learning networks performed well at high computational costs, heavy model structures and sufficient availability of labelled data for training. This led the researchers to explore solutions such as factorizing convolutions and dimensionality reductions to reduce computational costs of deep learning [39]. Mobile networks were designed to maintain classification accuracy in resource constrained environments [50]. Computational costs were also lowered using depthwise separable convolutional layers methodology [51]. The deep learning architectures were made more sparse [52] through network pruning [53–58], connectivity learning [59,60] and hyper parameter optimization [61–63]. In hyper parameter optimization, the deep neural networks were tuned using various algorithms [64] to enhance accuracy and performance at minimal computational costs. Newer methodologies such as reinforcement learning enabled optimum network architecture designs [65–68]. On the other hand, some researchers reported that heavily optimized deep networks prove to be too complex to be implemented in case of smaller training datasets and low resource applications [50]. Hence, transfer learning methodology was explored to successfully apply deep learning models in such cases [69]. The next subsection introduces transfer learning and its applicability aspects.

1.4. Classification using transfer learning

Transfer learning involves applying a network pre-trained on a large dataset to classify the features from a smaller dataset [70–72]. This approach results in higher accuracy as compared to directly training deep learning models on the smaller datasets for classification without pretraining [73], [74]. Hence, transfer learning is especially applicable for classification problems involving training datasets insufficient for deep learning implementation [75]. Recent applications of transfer learning include industrial fault diagnoses [76] such as injection moulding [77], turbine blades [78], wind power prediction [79] and TIG welding defects [69]. Advanced tuning and optimization techniques can further improve the prediction/classification accuracy of transfer learning methodology [80]. The following section furnishes details of the experimentation and transfer learning methodology carried out in the current study.

2. Transfer learning methodology

As explained in the previous section, transfer learning refers to the training based classification of target dataset using deep networks that have already been pretrained on a much larger dataset. The transfer learning methodology adopted in the current study involved the following steps -

1. Data collection of acoustic emissions generated during turning of mild steel specimens
2. Generation of spectrograms for all acoustic data samples using mel frequency cepstral coefficient (MFCC) and sample rate filter
3. Building transfer learning framework using -
    (a) Pre-trained deep learning models as layer 0
    (b) Fully connected network as the subsequent layer
4. Training the transfer learning architecture on the preprocessed spectrogram images for multi class classifications
5. Maximizing model performances by using suitable optimizers

2.1. Experimentation: Acoustic data collection

The machining experiments were conducted on a CNC turning centre (make: LMW machining, model: LX20TL3) using coated carbide insert (make: Duracarb, code: TNMG 160408 43 DC9025) as cutting tool and cylindrical rods of mild steel (diameter 30 mm and length 100 mm each) as raw material. Figs. 1 and 2 show the CNC turning centre and the cutting tool insert (respectively) used in the current study.

The three edges of the cutting tool insert were utilized to machine three different workpieces of the same material under varying cutting speed levels to generate progressively variable tool wear on them (Table 1). The insert edge A was used to machine mild steel for consecutive ten runs with cutting speeds varied from 70 to 700 rpm in equal increments. Insert edge B machined mild steel consecutively for twenty runs from 70 to 1400 rpm in equal steps. Finally, insert edge C machined mild steel rods consecutively for fifty runs from 70 to 3500 rpm in equally spaced increments. Each experimental run was executed for a uniform period of ten seconds. Thus, insert edge B was used to machine for twice as many experimental runs with progressively increasing cutting speeds as compared to insert edge A. Similarly, the insert edge C was employed to machine much greater number of experimental runs

Fig. 1. CNC Turning Center.

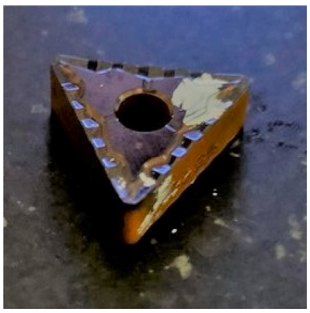


Fig. 2. Cutting Tool Insert.

as compared to edges A and B. In total, eighty experiments were carried out. This experimental design ensured generation of three types of progressive tool wear on these three edges of the cutting tool insert. Insert edge A tool wear (at the end of ten experimental runs) was labelled as low, whereas the tool wear generated on the flank face of insert edges B and C were labelled medium and high respectively (at the end of twenty and fifty experimental runs respectively). The feed rate and the depth of cut were kept constant in all experiments at 0.15 mm/rev and 0.5 mm respectively. Fig. 3(a) shows the toolmaker's microscope used to take pictures of the tool flank wear land for insert edges A, B and C. Fig. 3(b) shows how the cutting tool insert edge was positioned below the microscopic lens for capturing the wear images.

The machining noise generated during each experimental run was recorded using homogeneous smartphone microphone sensors. One sensor was placed close to the CNC turning centre whereas the other was placed at a distance of ten feet from the machine for a period of seven seconds. Thus, a total of 160 machining audio samples (of 7 s each) were collected from 80 experiments. The initial ten experiments involved machining at low speeds, producing low tool wear. Hence, the acoustic emission recordings corresponding to experimental runs 1 to 10 (Table 1) were labelled as 'low wear' for all inserts. In this way, 30 low wear recordings were obtained, 10 each from insert edges A, B and C. The experimental runs 11 to 20 involved insert edges B and C consecutively machining at progressively higher cutting speeds, generating medium wear on the tool flank face. Therefore, the sonic recordings from experimental runs 11 to 20 were labelled 'medium wear' from both insert edges B and C, yielding 10 samples each and 20 samples in total. Finally, the machining noises recorded during the experiments 21 to 50 for insert edge C yielded 30 samples labelled as 'high wear' class, due to the progressively higher cutting speeds involved in the consecutive experiments in this range. All machining noise data was initially collected in mp3 format and later converted to .wav for subsequent filtering and analysis. The following section details upon the analysis carried out on the tool wear machining acoustic data.

2.2. Acoustic data preprocessing

After the sonic emissions were recorded under different machining conditions, data features were extracted from the recordings (acoustic emission) using spectrograms generated using mel frequency cepstral coefficient (MFCC) and sample rate filters. MFCC

(a) (b)

Fig. 3. Tool Makers Microscope.

spectrograms have been found useful by researchers in varied audio classification applications ranging from UAVs [81] to home appliances [82]. The MFCC and the sample rate filters extract intricate features which help differentiate signals of different wavelengths and frequencies. Such features have a major role in representing the unique characteristics of each acoustic emission by which it can be classified as per the tool wear condition at which it was recorded. Spectrograms depict the features and frequencies of the acoustic signals in the form of two or three dimensional visual representations that are easier to classify as compared to one dimensional audio signals. In the present study, two dimensional spectrograms were derived from MFCC and sample rate filters using Matlab. Firstly, the .wav sound files were filtered using Butterworth sample rate filter and MFCC codes in Matlab. Thereafter, the filtered audio signals were used to plot spectrograms using Matlab codes made available by Zhivomirov, titled 'Sound Analysis with Matlab Implementation' in the Matlab Central File Exchange (version 1.4.0.0) [83].

Fig. 4 shows the tool flank face images captured by the tool maker's microscope for each of the insert edges A, B and C after the completion of respective experimental trials. Fig. 5 shows sample spectrograms for each of the three data labels considered in the acoustic recordings: low, medium and high tool wear. The left and right y-axes of these figures show the audio signal frequencies and decibels respectively. Furthermore, the spectrograms shown in Fig. 5(a), (b) and (c) correspond to A1 (insert edge A, cutting speed 70 rpm, feed rate 0.15 mm/rev, depth of cut 0.5 mm), B20 (insert edge B, cutting speed 1400 rpm, feed rate 0.15 mm/rev, depth of cut 0.5 mm) and C50 (insert edge C, cutting speed 3500 rpm, feed rate 0.15 mm/rev, depth of cut 0.5 mm) experimental runs, respectively.

Hence, a total of 160 spectrogram images were derived from acoustic emissions from the same number of experimental trials. All these spectrogram images were distributed among training (80 %), validation (10 %) and testing (10 %) data subsets for modelling deep learning models. The above mentioned distribution of data samples into training, testing and validation datasets is usually recommended in machine learning/deep learning research literature [84]. Maximum samples are reserved for training the algorithms, followed by smaller sample sizes reserved for validation and testing tasks. The algorithms are trained against the training datasets and validated against the validation datasets at the end of every training epoch. Finally, the fully trained and validated algorithms are tested to perform against an unseen testing sample set.

Table 2 shows the number of spectrogram images belonging to the three tool wear classes. Of the total 60 spectrogram images available for each of the low and high tool wear classes, 48 were allotted for training whereas 6 each were reserved for validation and testing. Similarly, the 40 medium wear class spectrogram pictures were distributed as 32, 4 and 4 for training, validation and testing respectively. The training datasets of spectrogram images of respective wear classes were subsequently utilized for classification by the deep learning models presented in the following subsection.

Table 1
Experimental design.

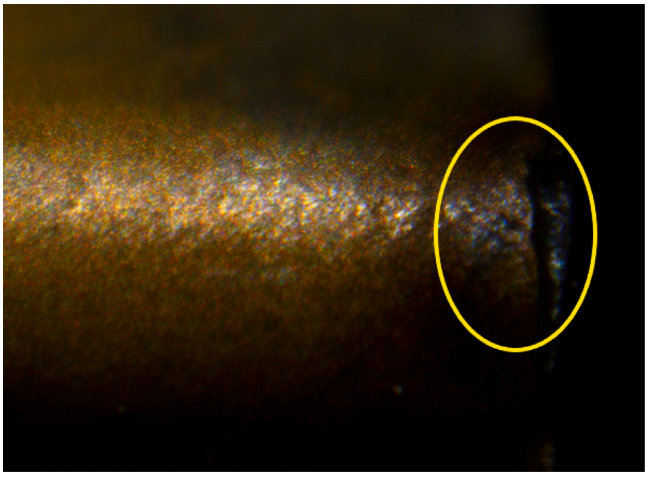
Sr. no.	Cutting speed (rpm)	Feed rate (mm/rev)	Depth of cut (mm)	Tool insert edge		
1	70	0.15	0.5	A	B	C
2	140	0.15	0.5	A	B	C
3	210	0.15	0.5	A	B	C
4	280	0.15	0.5	A	B	C
5	350	0.15	0.5	A	B	C
6	420	0.15	0.5	A	B	C
7	490	0.15	0.5	A	B	C
8	560	0.15	0.5	A	B	C
9	630	0.15	0.5	A	B	C
10	700	0.15	0.5	A	B	C
11	770	0.15	0.5		B	C
12	840	0.15	0.5		B	C
13	910	0.15	0.5		B	C
14	980	0.15	0.5		B	C
15	1050	0.15	0.5		B	C
16	1120	0.15	0.5		B	C
17	1190	0.15	0.5		B	C
18	1260	0.15	0.5		B	C
19	1330	0.15	0.5		B	C
20	1400	0.15	0.5		B	C
21	1470	0.15	0.5			C
22	1540	0.15	0.5			C
23	1610	0.15	0.5			C
24	1680	0.15	0.5			C
25	1750	0.15	0.5			C
26	1820	0.15	0.5			C
27	1890	0.15	0.5			C
28	1960	0.15	0.5			C
29	2030	0.15	0.5			C
30	2100	0.15	0.5			C
31	2170	0.15	0.5			C
32	2240	0.15	0.5			C
33	2310	0.15	0.5			C
34	2380	0.15	0.5			C
35	2450	0.15	0.5			C
36	2520	0.15	0.5			C
37	2590	0.15	0.5			C
38	2660	0.15	0.5			C
39	2730	0.15	0.5			C
40	2800	0.15	0.5			C
41	2870	0.15	0.5			C
42	2940	0.15	0.5			C
43	3010	0.15	0.5			C
44	3080	0.15	0.5			C
45	3150	0.15	0.5			C
46	3220	0.15	0.5			C
47	3290	0.15	0.5			C
48	3360	0.15	0.5			C
49	3430	0.15	0.5			C
50	3500	0.15	0.5			C

Table 2
Data distribution of machining dataset for acoustic classification.

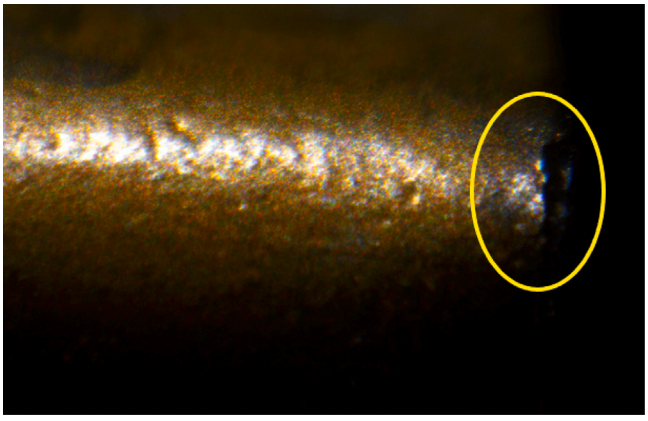
Class	Total samples	Training (80%)	Validation (10%)	Testing (10%)
Low	60	48	6	6
Medium	40	32	4	4
High	60	48	6	6
Total	**160**	**128**	**16**	**16**

2.3. Deep learning models and optimizers

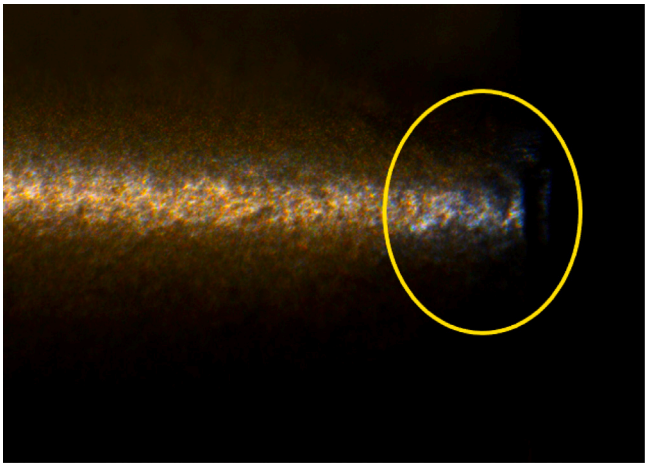
This subsection firstly introduces the pre-trained deep learning models used in the current study. Later, it briefly introduces the optimizers employed to boost performances of the deep learning models used in the current study.



(a) Edge A: low wear



(b) Edge B: medium wear

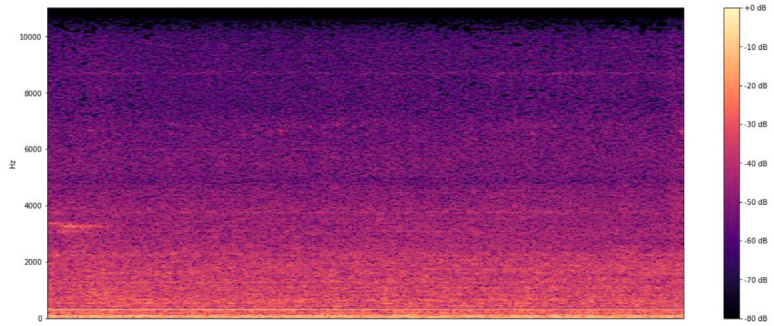


(c) Edge C: high wear

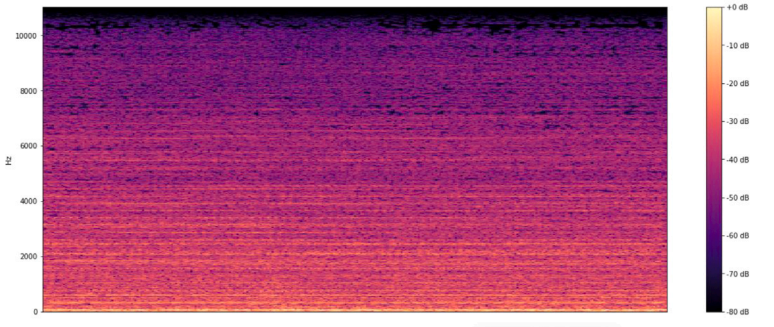
Fig. 4. Tool flank wear on insert edges.

2.3.1. SqueezeNet

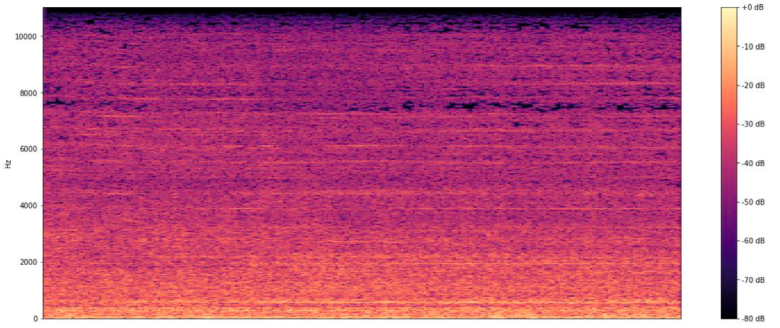
SqueezeNet is a type of convolutional neural network (CNN) that 'squeezes' CNN layers to yield a much smaller, compact and effective architecture ideally suited for relatively low resource embedded applications such as mobile phones and FPGAs [85].



(a) Low wear: experiment A1



(b) Medium wear: experiment B20



(c) High wear: experiment C50

Fig. 5. Sample spectrograms.

SqueezeNet is able to generate classification accuracy equivalent to that of any standard CNN, at the cost of much fewer parameters [86]. This network can be optimized during training for optimum results and has been applied in several diverse domains [87].

2.3.2. GoogLeNet

GoogLeNet architecture was developed in 2014 for object detection and image classification tasks [42]. It is composed of 22 layers and inception modules to execute multiple convolutional transformations for feature extraction. It aims to maximize accuracy at minimal computational costs. Its drawbacks included representation jam due to heterogeneous topology, which might lead to loss of some valuable information/data features [88].

2.3.3. ResNet50

Residual networks (ResNets) were developed in 2016 to simplify the deep neural network training process [89] by learning through the residual functions present in multiple layers. These residual networks were easy to optimize and their accuracy could be

Table 3
Feature map sizes in pre-trained deep learning models.

Sr. No.	Model	Feature map size
1	SqueezeNet	512
2	GoogLeNet	1024
3	ResNet50	2048
4	InceptionV3	2048

improved at greater layer depths. Researchers [89] claimed the superior performance of the 50 layered residual network (ResNet50) on the ImageNet classification problem over many others.

2.3.4. InceptionV3

InceptionV1 was introduced in 2014 [42] and its improved versions InceptionV2 [90] and InceptionV3 [91] were developed in the subsequent years. These models have been shown to perform exceedingly well on the ImageNet dataset [92]. Inception architectures are composed of multiple convolutional feature extractors called inception modules. These modules are designed to involve fewer parameters for learning data patterns. Hence, inception models tend to consume lesser computational costs and are well suited for constrained memory applications [93]. However, in some cases inception models have been found to suffer from inflexibility and complexity that can be handled by suitable optimization.

The above mentioned deep learning models were optimized by three optimizers introduced as follows.

2.3.5. SGDM

The stochastic gradient descent (SGD) optimizer was developed to maximize learning while minimizing computational complexities during training [94]. SGD consumes minimal training time since it considers only a single random data sample to update its gradients over training epochs. SGDM is SGD with a degree of 'momentum' accorded to its gradient vectors to converge faster towards the right direction.

2.3.6. ADAM

The Adaptive Moment Estimation (ADAM) is an optimization algorithm used for computing optimal solutions in neural networks [95]. Adam uses exponential decay rates (β_1 and β_2) learning rate (g_t) and to determine step sizes during back propagation for convergence.

2.3.7. RMSPROP

The root mean squared propagation (RMSPROP) algorithm was developed to overcome the zero learning rate limitation of the ADAGRAD method. The accumulation of gradients over successive iterations sometimes causes ADAGRAD's learning rate to became zero. RMSPROP overcomes this limitation by firstly computing the moving average of the squares of all preceding gradients. Then, it divides the current gradients by the square root of this moving average to avoid the occurrence of zero learning rate.

The above mentioned deep learning models and optimizers were chosen on the basis of a previous study carried out by the authors, in which similar multiple model-optimizer combinations were explored that proved useful in maximizing the classification accuracy of welding defects [96]. The following subsection gives details of how the above discussed deep learning models were utilized in a transfer learning architecture for three-class tool wear classification. The subsection also explains the various metrics used to evaluate the performance of different model-optimizers.

2.4. Transfer learning architecture and performance metrics

Fig. 6 depicts the process flow of the transfer learning methodology adopted in the current study. The preprocessed acoustic spectrogram images were input to the three class transfer learning architecture, wherein the layer 0 consisted of the pre-trained deep learning models. The subsequent layers of the transfer learning architectures were formed by the first and second fully connected layers (FC1, FC2) having 256 and 64 neurons respectively. To avoid overfitting of the model, a dropout rate of 0.5 was considered in FC2, which was followed by the output layer producing the three-class tool wear classifications. Table 3 depicts the feature map sizes (sizes of the second last layers) of the four pretrained deep learning models employed in the current work, viz. SqueezeNet, GoogleLeNet, ResNet50 and InceptionV3. These pre-trained deep learning architectures were further tuned using three optimizers — SGDM, ADAM and RMSPROP. Thus, a total of 12 different combinations of deep learning models and optimizers were investigated for transfer learning of three-class tool wear categorizations.

The performances of all deep learning models were evaluated using the standard metrics followed in literature [97] - testing accuracy, precision, recall and F1 scores. Prediction accuracy is defined as follows

$$Accuracy = \frac{Correct\ predictions}{Total\ samples} \tag{1}$$

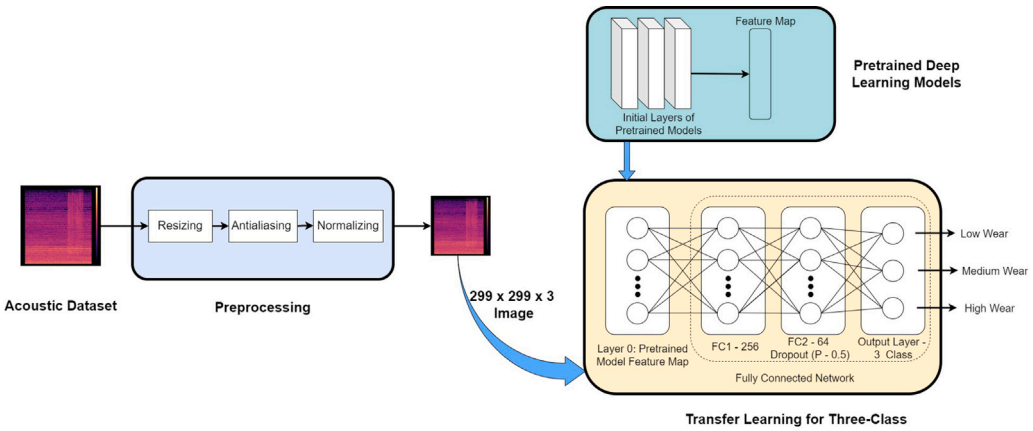


Fig. 6. The flowchart of transfer learning methodology for acoustic wear classification.

The above definition of accuracy compares the sum of true negatives (TN) and true positives (TP) against the sum of false positives (FP), false negatives (FN), true negatives (TN) and true positives (TP). Model precision is used to express prediction accuracy of the predicted samples, determined as follows

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

On the other hand, recall is used to express prediction accuracy over the total number of samples, given as

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

High precision of high tool wear classification is desirable to ensure that minimum good tools are wrongly categorized as totally worn out. This factor minimizes the production pauses for unnecessary tool replacements. On the other hand, high recall value of high tool wear classification ensures that minimum number of highly worn out tools are incorrectly classified as still good to machine. This factor is critical to ensure high production quality. Similarly, high precision of medium and low wear classifications ensure that minimum badly worn tools are incorrectly classified as okay. Lastly, high recalls of low and medium wear classifications minimize good tools being misidentified as bad and thus being unnecessarily replaced. Given the importance of both precision and recall, F1 scores are used to obtain a balanced measure of both metrics. F1 scores are able to handle unbalanced datasets effectively and are expressed as follows

$$F1\ Score = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}} = \frac{2 * Precision * Recall}{(Precision + Recall)} \tag{4}$$

The optimizer performances were scrutinized through the training loss/accuracy evolutions over successive epochs for each of the deep learning models considered in the current study. These evolutions primarily indicate the degrees of over/underfitting of a model to the training datasets. If the training loss is much lesser than validation loss, it indicates an overfit model that is not generalized enough to handle unseen data. On the other hand, if higher training loss shows that the model is underfitting the training dataset substantially. The comparable magnitudes of validation/training losses and their convergence over successive epochs confirm a well fitted deep learning architecture that can handle unseen data properly. The following section presents the results and associated discussions of the proposed tool wear classification methodology.

3. Results and discussions

In the present study, transfer learning methodology was applied for acoustic data based tool wear classification. Pre trained deep learning models were utilized to classify the acoustic records into multi class classifications viz. high, medium and low tool wear regimes. In the present study four different deep learning architectures were explored, in association with three optimizers each, resulting in a total of 12 model-optimizer configurations.

3.1. Based on testing accuracy

Table 5 depicts the testing accuracies of the twelve optimizer-model combinations explored in the current study. These testing accuracies quantify the ability of the explored model architectures to correctly identify tool wear modes - low, medium and high. It is evident from the Table 5 that the Inception v3 model optimized by RMSPROP attained the highest testing accuracy of 87.50 %, followed by the RMSPROP optimized SqueezeNet model having 75.00 % accuracy. The SGDM optimized ResNet50 model achieved

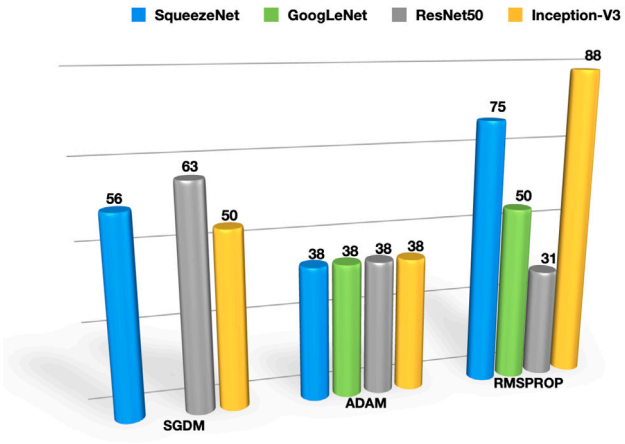


Fig. 7. Three class testing accuracies (%) tool wear classifications.

the third best testing accuracy of 62.50 %. Except for ResNet50, the RMSPROP maximized testing accuracies of all models explored in the current study. In contrast, the ADAM optimized model performances were far inferior. In the case of ResNet50, the SGDM maximized classification accuracy. On the other hand, the RMSPROP fared poorly in comparison (for ResNet50). It is also to be noted that the SGDM optimized GoogLeNet model was unable to successfully classify the training data considered in the present work. Fig. 7 depicts the above discussed multi-class accuracies of all model-optimizer architectures pictorially.

In general, RMSPROP outperforms SGDM due to its adjustable learning rate, stabilized learning algorithm and fewer hyper parameters to be tuned as compared to SGDM. On the other hand, SGDM performs better than ADAM (another adaptive optimizer) across all models, due to its better generalization capabilities. ADAM generally works better than SGDM due to its adaptive learning rates and faster convergence. However, in case of sparse datasets (such as the one used in the present study), it tends to become more sensitive to noise elements in the dataset, leading to divergence or subopotimal convergence (as evident in the present study). Hence, these results imply that in general, while using large datasets, adaptive learning rate optimizers such as RMSPROP and ADAM can be expected to obtain better optimized solutions in lesser convergence timescales (although at higher computational costs). However, in case of handling sparser datasets, ADAM is unable to optimize the deep learning models optimally, falling short of even standard learning rate optimizers such as the SGDM.

Regarding the performance of deep learning models, results show that ResNet50 (which is a 50-layer deep architecture) works better with a standard learning rate optimizer (SGDM) as compared to adaptive learning rate optimizers in case of sparse datasets. On the other hand, Inception V3 is capable of reducing its parameter set by decomposing large layers into smaller convolutional layers, making it suitable to handle small/sparse datasets. SqueezeNet also boasts of a similar capability to 'squeeze' its parameters, and is thus suitable for smaller datasets. However, it is only 18 layers deep as compared to the 48 layers of Inception V3, which make it relatively inferior to Inception V3's modelling capabilities. GoogLeNet has only 22 layers with a single inception module, and is evidently unable to handle complexities of modelling smaller datasets as effectively as other models such as Inception V3 and even SqueezeNet. The above discussed results imply that in case of smaller datasets, the deep learning models with the capability of collapsing/squeezing larger layers/ greater parameters into smaller layers/ lesser parameters must be employed in conjunction with an adaptive rate equipped and stable learning optimizer (such as RMSPROP) for better results in future research endeavours.

The following section presents a more detailed analysis of optimizer performances across the four deep learning models considered in the current study.

3.2. Based on precision, recall and F1 scores

Table 4 displays the three class classification precision, recall and F1 scores for all twelve model-optimizer combinations explored in the current study. The following subsections present an analysis of these results from different perspectives. Firstly, overall performance comparisons are drawn between the model-optimizer having the best testing accuracy (InceptionV3-RMSPROP) and all others. The second and third subsections discuss a physical machining based analysis of the precision and recall performances (respectively) of the top three model-optimizers having the best testing accuracies (InceptionV3-RMSPROP, SqueezeNet-RMSPROP and Resnet50-SGDM). Fourthly, a similar performance analyses of the remaining architectures of the better performing models (considering testing accuracies) — SqueezeNet and InceptionV3 is presented. The fifth subsection presents a similar analysis of the relatively lower testing performance models of ResNet50 and GoogLeNet. The sixth subsection compares the relatively lower testing performing optimizers — ADAM and SGDM across all models, followed by the seventh and the last subsection that discusses the effect of dataset sizes on the precision, recall and F1 scores.

Table 4
Performance evaluation based on precision, recall and F1 scores for acoustic classification.

		Precision			Recall			F1-Score		
Sr. No.	Model - Optimizer	H	M	L	H	M	L	H	M	L
1	SqueezeNet - SGDM	0.50	0.00	0.75	1.00	0.00	0.50	0.67	0.00	0.60
2	SqueezeNet - ADAM	0.38	0.00	0.00	1.00	0.00	0.00	0.55	0.00	0.00
3	SqueezeNet - RMSPROP	1.00	0.50	1.00	0.83	1.00	0.50	0.91	0.67	0.67
4	GoogLeNet - SGDM	No classification								
5	GoogLeNet - ADAM	0.38	0.00	0.00	1.00	0.00	0.00	0.55	0.00	0.00
6	GoogLeNet - RMSPROP	0.00	0.33	1.00	0.00	1.00	0.67	0.00	0.50	0.80
7	ResNet50 - SGDM	1.00	0.00	0.50	0.67	0.00	1.00	0.80	0.00	0.67
8	ResNet50 - ADAM	0.00	0.00	0.38	0.00	0.00	1.00	0.00	0.00	0.55
9	ResNet50 - RMSPROP	0.00	0.27	1.00	0.00	1.00	0.17	0.00	0.42	0.29
10	Inception-V3 - SGDM	0.42	0.00	0.75	0.83	0.00	0.50	0.56	0.00	0.60
11	Inception-V3 - ADAM	0.00	0.00	0.38	0.00	0.00	1.00	0.00	0.00	0.55
12	Inception-V3 - RMSPROP	1.00	0.67	1.00	0.67	1.00	1.00	0.80	0.80	1.00

3.2.1. Performance comparison of the top model-optimizer (having the highest testing accuracy) vs. others

As discussed in the previous section, the InceptionV3-RMSPROP architecture attained the best testing accuracy of 87.50 %. This performance is duly reflected in its superior values of precision, recall and F1 values over all other model-optimizer combinations except in case of recall and F1 scores of high tool wear class wherein some of the other model-optimizer architectures performed better. For instance, the SqueezeNet model optimized by all three optimizers viz. SGDM, ADAM and RMSPROP attained better recall scores (1.00, 1.00 and 0.83 respectively) for high tool wear class as compared to that of the Inception V3 - RMSPROP (0.67). The GoogLeNet - ADAM and Inception V3 - SGDM also obtained better recall scores (1.00 and 0.83) for high tool wear class classification over that of the Inception V3 RMSPROP architecture. However, in case of the F1 scores only the SqueezeNet - RMSPROP model-optimizer combination attained better output (0.91) as compared to that of the Inception V3 RMSPROP (0.80), that too only in case of the high tool wear classification.

3.2.2. Physical implications of the precision performances of the top three model-optimizers (with the highest testing accuracies)

Considering the precision performance of the top two model-optimizer pairs with the highest testing accuracies viz. Inception V3 - RMSPROP and SqueezeNet - RMSPROP (Table 5), it is evident that both of them attained the highest possible precision for high and low tool wear classes i.e. 1.0. Perfect precision means that there were zero instances of false positive classifications for these classes. In other words, these two model-optimizer pairs did not make any mistake of misidentifying a low or medium tool wear acoustic as a high tool wear. Similarly, they did not misidentify any high or medium tool wear acoustic signature as an instance of low tool wear. From a quality and safety perspective, it is better to have perfect precision of medium and low tool wear classifications as compared to that of the high tool wear class. This is because during production, the manufacturing firm cannot risk continue using a worn tool misidentified as having low or medium tool wear. Continued usage of a highly worn tool will deteriorate the quality of the machined components and lead to their rework or rejection during the subsequent quality inspections. In extreme cases, such highly worn tools may damage the machine and/or cause operator injury as well. Hence, perfect precision of low and medium tool wear classifications is highly desirable. On the other hand, imperfect/inaccurate precision of high tool wear classification will imply that some of the low/medium worn tools are misidentified as highly worn. This imperfect classification will not hurt the enterprise severely, since it will only lead to higher cutting tool replacements. The cost of cutting tool replacements is almost always lower than that of rework/part rejection/safety hazards.

Nevertheless, in the current study, both of the top performing model-optimizer pairs attained imperfect/inaccurate precision for medium tool wear mode classes. This indicates the presence of false positives in medium tool wear identification. In other words, some high or low worn tool acoustics were misidentified as those of medium worn tools. The misidentification of low worn tools as medium is tolerable, but misclassification of highly worn tools as medium is detrimental. Hence, the RMSPROP optimized Inception V3 model is relatively more acceptable over the SqueezeNet - RMSPROP due to its superior precision (0.67) of medium tool wear classification over the latter (0.50).

The third best optimized model (as per testing accuracy) i.e. ResNet50 - SGDM achieved perfect precision for high wear (1.0) but zero and 0.5 precision for medium and low tool wear modes respectively. As per the above discussed practical machining implications, the performance of this model-optimizer is relatively undesirable.

3.2.3. Physical implications of the recall performances of the top three model-optimizers (with the highest testing accuracies)

Next, considering the recall accuracies of the Inception V3 - RMSPROP and SqueezeNet - RMSPROP, both achieved perfect recall for medium tool wear. Inception V3 - RMSPROP also got perfect recall for low tool wear but SqueezeNet - RMSPROP achieved superior recall for the high wear classification. Perfect recall implies zero instances of false negatives in classification. In other words, these two model-optimizer pairs made no mistake of misidentifying an actual medium tool wear acoustic signature as

something else (high/low). Similarly, the Inception V3 - RMSPROP did not misidentify any genuine low worn tool as medium or highly worn. Finally, the SqueezeNet - RMSPROP misidentified lesser high worn tools as low/medium wear as compared to such misclassifications by the Inception V3 - RMSPROP. From a practical machining point of view, misidentification of a low/medium worn tool as something else is not a severe mistake. Hence, slightly imperfect recall values are tolerable in case of low/medium worn tools. However, misclassification of a highly worn tool as otherwise is unacceptable. Highest possible recall values are desirable in this case. The SqueezeNet - RMSPROP outperforms Inception V3 - RMSPROP in this regard and is more acceptable inspite of having lesser recall accuracy (0.50) in low tool wear class as compared to the perfect recall attained by the latter architecture.

However, the foregoing discussion on precision metrics of top performing model-optimizers concluded that the InceptionV3-RMSPROP was better than the SqueezeNet-RMSPROP owing to the slightly better medium wear class precision of the former over the latter. Therefore, further discussion is required as to which of these two model-optimizers is more preferable for acoustics based tool wear classification. It may be noted that lack of precision in medium wear class implies false positives that may include both high and low tool wear cases misidentified as medium wear; whereas lack of recall in high wear implies some false negatives which definitely include high tool wear instances falsely classified as low or medium. Hence, recall of high tool wear is more critical as compared to precision of medium wear class. Moreover, deficiency in the high tool wear recall (0.83) of SqueezeNet-RMSPROP is lesser than that in the medium wear precision (0.67) of InceptionV3-RMSPROP. Thirdly, the high tool wear F1 score (0.91) of SqueeNet-RMSPROP is superior to that of InceptionV3-RMSPROP (0.80). Therefore, it may be definitively concluded that the SqueezeNet-RMSPROP is more suitable for acoustic based tool wear mode classification in machining in spite of having lesser testing accuracy (75.00 %) than InceptionV3-RMSPROP (testing accuracy 87.50 %).

The third best optimized model (as per testing accuracy) i.e. ResNet50 - SGDM achieved perfect recall for low wear (1.0) but zero and 0.67 precision for medium and high tool wear modes respectively. As concluded regarding its precision performance, the recall attainments of this model-optimizer are also undesirable from the practical machining viewpoint.

3.2.4. SqueezeNet versus InceptionV3 performances across SGDM and ADAM optimizers: physical implications

On comparing the performances of the SqueezeNet and InceptionV3 models optimized by SGDM, results show that both these model-optimizers obtained 0.00 and 0.75 precision for the medium and low tool wear classes respectively. The SqueezeNet-SGDM obtained a slightly better high tool wear class precision of 0.50 over that of its IncpetionV3-SGDM counterpart (0.42). Considering the greater importance of the medium and low wear class precisions over that of the high wear mode from a physical machining perspective, both these model-optimizers stood equal in precision performances, although these performances were not impressive from a physical machining viewpoint. The recall performances of both these model-optimizers were also exactly equal for the medium (0.00) and the low (0.50) wear classes. However, the SqueezeNet-SGDM attained a perfect recall (1.00) for the high tool wear class over that of the InceptionV3-SGDM (0.83). Given the highest importance accorded to the high tool wear class recall, it may be concluded that the SqueezeNet-SGDM outperformed InceptionV3-SGDM. Similarly, the SqueezeNet-ADAM outperformed the InceptionV3-ADAM based on the perfect high wear class recall (1.0) of the former as compared to the 0.00 high wear recall value of the latter model-optimizer. Among the SGDM and ADAM optimized SqueezeNet models, the former outperforms the latter based on its higher low tool wear precision.

3.2.5. GoogLeNet versus ResNet50 performances across all optimizers: physical implications

On comparing the performances of GoogLeNet and ResNet50 models optimized by ADAM, the results show that the former architecture achieved 0.38 precision for high tool wear and 0.00 precision for low and medium wear classes. Whereas, the latter combination achieved 0.38 precision for low tool wear and 0.00 precision for high and medium classes. In the light of the aforementioned physical machining viewpoint, the ResNet50-ADAM's precision performance is more desirable than that of GoogLeNet-ADAM. In terms of recall, the GoogLeNet-ADAM achieved perfect score of 1.00 for high wear class and 0.00 score for medium and low classes. The ResNet50-ADAM on the other hand, scored perfect 1.00 for low wear class and 0.00 for medium and high wear. From practical machining viewpoint, GoogLeNet-ADAM's recall performance is more desirable. Furthermore, since the importance of high wear recall is greater than the precision of low and medium wear classes, it may be concluded that the GoogLeNet-ADAM is more preferable over ResNet50-ADAM. The high wear recall of GoogLeNet-ADAM (1.00) is also quite superior to the low wear precision of ResNet50-ADAM (0.38). Moreover, the medium wear precision of ResNet50-ADAM is also 0.00, which weakens its case against the perfect high wear recall GoogLeNet-ADAM.

A comparison among the GoogLeNet and ResNet50 optimized by RMSPROP shows that both of these models obtained equal precision for the high wear (0.00) and low wear (1.00) classes. In terms of medium wear class precision, the GoogLeNet fared slightly better (0.33) over ResNet50 (0.27). In terms of recall, both models have equal values for the high wear (0.00) and medium tool wear (1.00) classes. For the low wear class, the GoogLeNet fared much better (0.67) relative to ResNet50 (0.17). Hence, the F1 scores of GoogLeNet-RMSPROP are superior to those of ResNet50-RMSPROP. However, the 0.00 high tool wear recall scores of both these model-optimizers is totally unacceptable from the physical machining perspective. Similar comparitive analyses cannot be conducted between the SGDM optimized GoogLeNet and ResNet50 models, since the former architecture was unable to produce any classifications for all tool wear categories explored in the current study.

3.2.6. ADAM and SGDM optimizer performances across models

Apart from the fact that no optimizer could enable models' classification of medium wear other than RMSPROP, the ADAM optimizer was unable to help ResNet50 and InceptionV3 properly classify high tool wear acoustic signatures as well. The precision, recall and F1 scores of ResNet50 — ADAM and InceptionV3 — ADAM are 0.00 for both high and medium tool wear regimes. Interestingly, the precision, recall and F1 scores of SqueezeNet — ADAM and GoogLeNet — ADAM were 0.00 for medium and low tool wear classes. No wonder, the ADAM optimized models attained the lowest three class classification testing accuracies (Table 5). In general, the SGDM optimized deep learning models performed better than ADAM optimized architectures, except for GoogLeNet, wherein the SGDM optimization could not result in any classification precision, recall or F1 score in any of the tool wear classes. In fact, the ADAM optimized InceptionV3 model attained better low tool wear recall (1.0) as compared to the model's SGDM optimized counterpart (0.50 recall). And, in case of ResNet50, both the SGDM and ADAM optimized versions maximized the low tool wear mode recall to 1.00.

3.2.7. Effect of dataset sizes on model-optimizer performances

It may be noted from Table 2 that there were relatively lesser labelled dataset available for medium wear class in the current study. This could be the reason behind comparatively lower precision, recall and F1 scores attained by almost all model-optimizers for this class. It is interesting to note that all models optimized by RMSPROP have been able to maximize recall accuracy of medium class to the perfect score. In fact, except for RMSPROP, no other optimizer could enable any of the models to attain non-zero precision accuracy. Similarly, only RMSPROP optimized models could attain non-zero F1 scores. It may be further observed that the F1 scores obtained by the RMSPROP optimized models for medium wear class are appreciable relative to those of other classes. Hence, it is clearly evident from the results that RMSPROP optimizer can be applied for successful deep learning classification of low volume data samples as well. Conversely, there were relatively greater samples of high and low wear generated by the experimental design followed in this study. Hence, almost all model-optimizers were able to classify their acoustic signatures with generally better precision, recall and F1 scores as compared to their medium wear counterparts. Hence, the top performing deep learning models attained reasonably good precision, recall and F1 scores in spite of unbalanced and very less number of data samples available for training.

The following sections discuss the estimation loss and accuracy trends of all deep learning model-optimizers over successive training cycles.

3.3. Training loss and accuracy of optimizers across models

This subsection presents an analysis of training loss and accuracy evolution over successive iterations for the optimizers across the deep learning models for the machining noise data generated in the current work. This section throws light on finer details of the optimizers' performances during the training of deep learning models considered in the present work.

3.3.1. SGDM optimized models

Fig. 8 depicts the SqueezeNet, GoogLeNet, ResNet50 and InceptionV3 training loss curves over successive iterations under SGDM optimizer. Training loss indicates the sum of estimation errors made by a model in a particular iteration. Training loss curves indicate how well the optimized deep learning models learn from their training losses and try to minimize them in the upcoming iterations. The InceptionV3 training loss remains almost constant from the very first iteration till the end. The SqueezeNet training loss quickly drops below two within the first few iterations and settles along InceptionV3 loss curve from the 20th iteration. On the other hand, SGDM finds it difficult to control the training loss for ResNet50 and GoogLeNet. In fact, the SGDM optimized GoogLeNet is unable to continue after the twelfth iteration and is terminated abruptly. ResNet50 loss curve is stabilized at the level of InceptionV3 after 45-50 iterations. From the training perspective, SGDM performs well with InceptionV3 and SqueezeNet models. However, Table 5 shows highest training accuracy achieved by ResNet50 among all SGDM optimized models. The answer to this lies in the training accuracy curves of the SGDM optimized models.

Fig. 9 shows training accuracy evolution of the deep learning models with the SGDM optimizer. Training accuracy is the percentage estimation accuracy of a model in a particular training iteration. Training accuracy curves show how well the optimized deep learning models learn from their past accuracies and try to improve them further in the forthcoming iterations. Over almost 90 iterations, the training accuracies of all models increase marginally except for GoogLeNet. The SGDM optimized models attain average to above average training accuracies. The ResNet50-SGDM managed to achieve the highest training accuracy among all models at the last iteration. The testing accuracies (Table 5) of these models are close to their respective final training accuracies shown in Fig. 9. The training losses of these models (Fig. 8) are also not at minimal levels. Since the training loss is a summation of errors and training accuracy is complementary to the percentage error, the above discussed results imply that the SGDM optimized deep learning models made average errors (based on training accuracy) on at least a minor part of the dataset (based on training losses) supplied for modelling in the current study.

3.3.2. ADAM optimized models

Fig. 10 shows that ADAM faces initial difficulty in training the deep learning models and consumes 20 iterations to stabilize loss at around one for InceptionV3, GoogLeNet and SqueezeNet. As in case of SGDM, the ResNet50 training loss is stabilized by ADAM

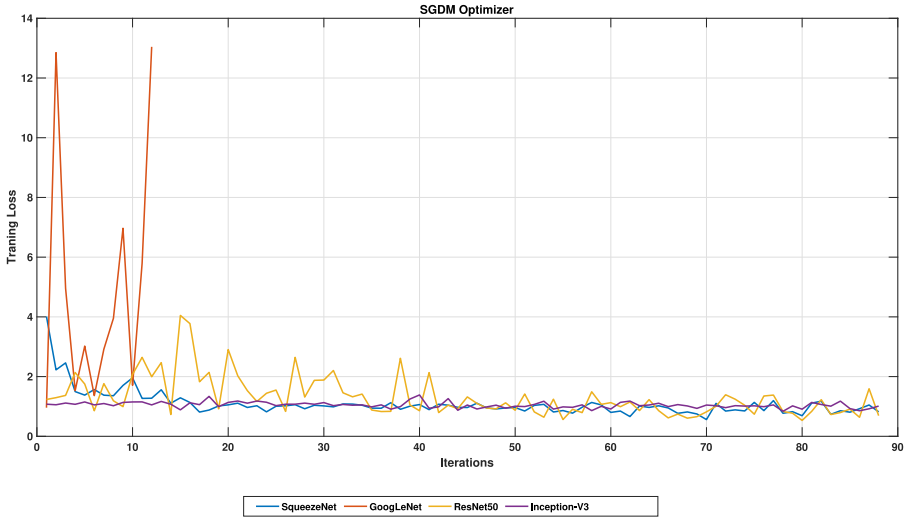


Fig. 8. Performance of training loss of SGDM optimizer across models for acoustic classification.

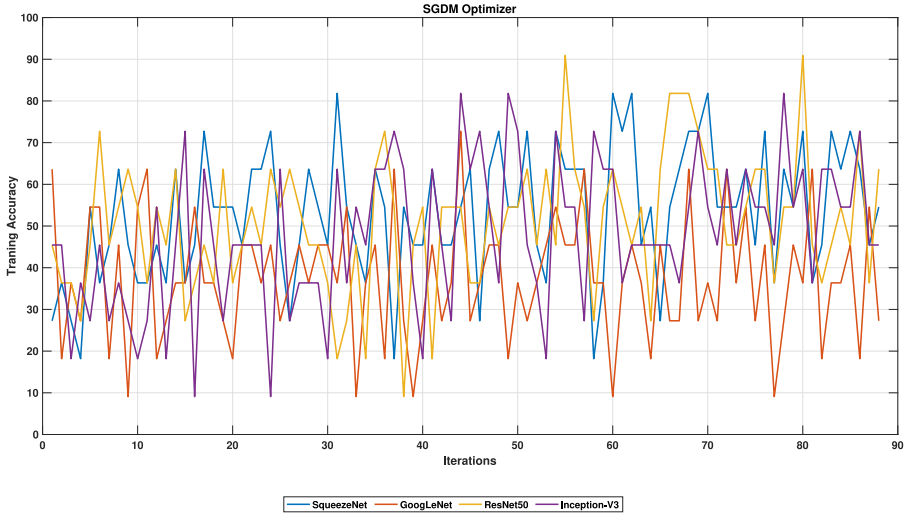


Fig. 9. Performance of training accuracy of SGDM optimizer across models for acoustic classification.

Table 5
Performance evaluation based on testing accuracies for multi-class classification (accuracy(%)).

Sr. No.	Model/Optimizer	SGDM	ADAM	RMSPROP
1	SqueezeNet	56.25	37.50	75.00
2	GoogLeNet	–	37.50	50.00
3	ResNet50	62.50	37.50	31.25
4	InceptionV3	50.00	37.50	87.50

after around 45 iterations. However, the final loss values of all models after almost 87 iterations are more or less similar to those attained by the SGDM optimizer. Still, the testing accuracies of SGDM optimized models are much superior to those optimized by ADAM (Table 5. The following discussion on training accuracy curves explains this phenomenon.

Fig. 11 shows training accuracy evolution of the deep learning models with the ADAM optimizer. Over almost 90 iterations, the training accuracies of all models do not record any increment except for GoogLeNet (which started at 10 % accuracy). The ADAM optimized models attain below average training accuracies. The testing accuracies (Table 5) of these models are close to their respective final training accuracies shown in Fig. 9, except from InceptionV3-ADAM, whose final training accuracy (less than 20 %) is much lesser than its testing accuracy (37.50 %). Similar to SGDM, he training losses of ADAM optimized models (Fig. 10)

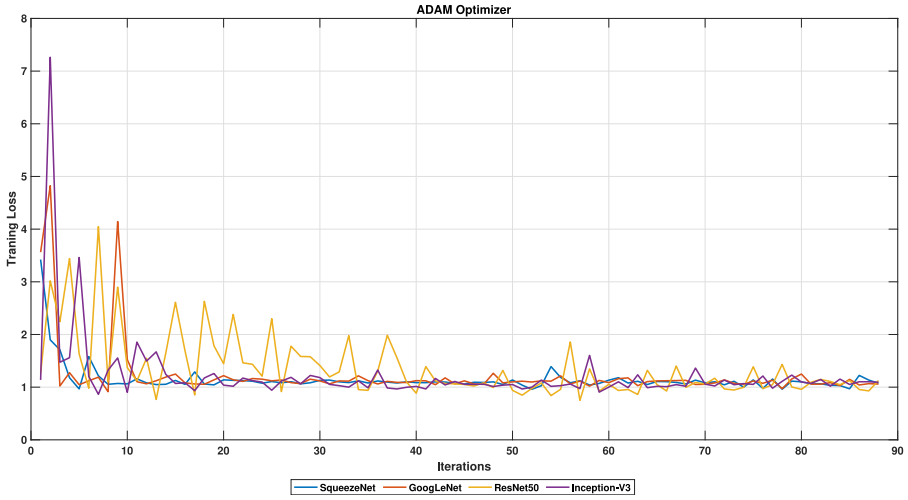


Fig. 10. Performance of training loss of ADAM optimizer across models for acoustic classification.

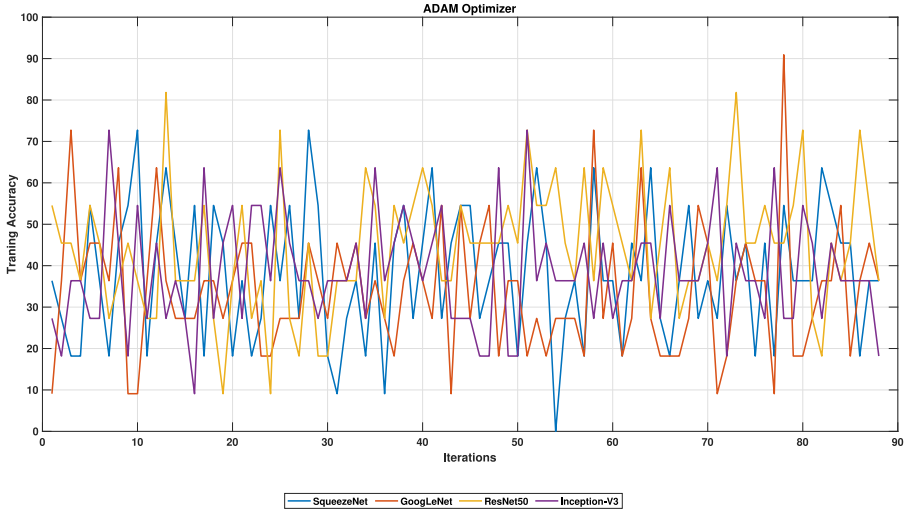


Fig. 11. Performance of training accuracy of ADAM optimizer across models for acoustic classification.

are also not at minimal levels. This implies that the ADAM optimized deep learning models made more than average errors (as per training accuracy) on a minor part of the dataset (as per training losses) supplied for modelling in the current study.

3.3.3. RMSPROP optimized models

Fig. 12 shows the RMSPROP is able to minimize training losses of all models below one by the end of all training iterations. This performance makes it superior to the above discussed SGDM and ADAM optimizers. Table 5 also depicts higher testing accuracies of RSMPROP optimized models over the others. In fact, RSMPROP is able to effectively limit the training loss of ResNet50 from the very beginning as compared to SGDM and ADAM.

Fig. 13 shows training accuracy evolution of the deep learning models with the RMSPROP optimizer. Over almost 90 iterations, the training accuracies of all models exhibit remarkable improvements. The InceptionV3, SqueezeNet, GoogLeNet and ResNet50 begin with less than 20 %, less than 30 % and around 45 % accuracies in the first iteration to more than 80 %, 60 %, 40 % and 80 % respectively in the final iteration. The testing accuracies (Table 5) of these models are even better than their respective final training accuracies shown in Fig. 9, except from ResNet50-RMSPROP, whose testing accuracy (31.25 %) is much lesser than its final training accuracy (above 80%). This implies overfitting of this model on the given dataset by the RMSPROP optimizer. Unlike SGDM and ADAM, the training losses of RMSPROP optimized models (Fig. 12) are below one. This implies that the RMSPROP optimized deep learning models made fewer errors (as per training accuracy) on a relatively smaller part of the dataset (as per the training losses) supplied for modelling in the current study.

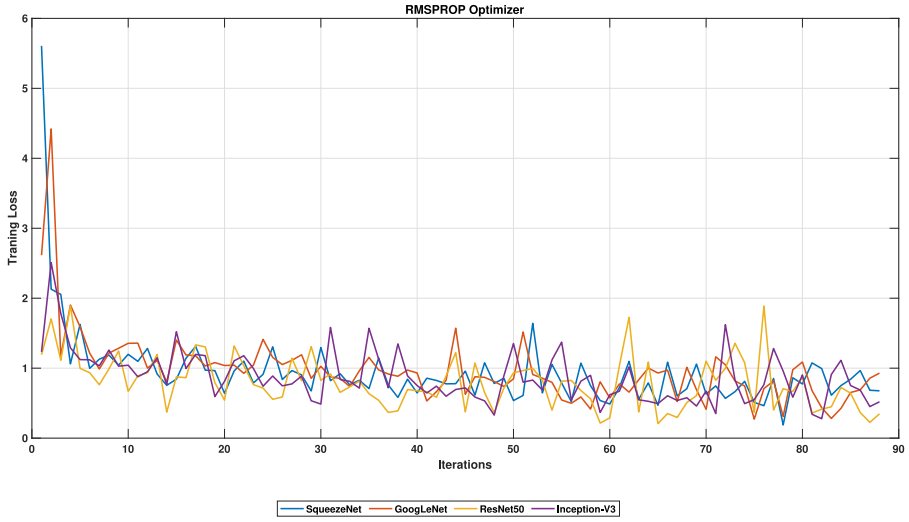


Fig. 12. Performance of training loss of RMSPROP optimizer across models for acoustic classification.

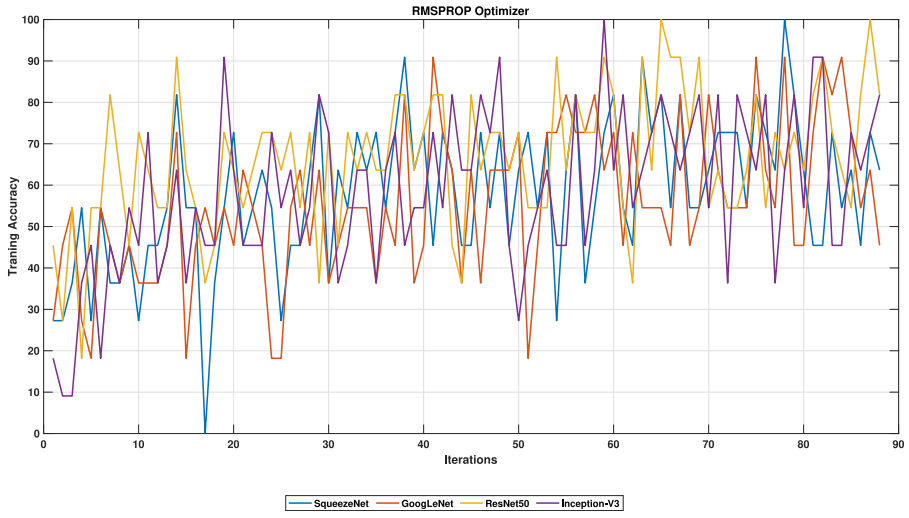


Fig. 13. Performance of training accuracy of RMSPROP optimizer across models for acoustic classification.

3.4. Training loss and accuracy of models across optimizers

This subsection presents an analysis of training loss and accuracy evolutions over successive iterations for the deep learning models across optimizers for the machining noise data generated in the current work. This section throws light on comparative performance of models under the different optimizers considered in the present work.

3.4.1. Optimizers in SqueezeNet model

Fig. 14 depicts the SGDM, ADAM and RMSPROP training loss curves over successive iterations for the SqueezeNet model architecture. As discussed in the previous section, training loss indicates the sum of estimation errors made by a model in a particular iteration. Training loss curves indicate how well the optimized deep learning models learn from their training losses and try to minimize them in the upcoming iterations. All optimizers bring down the training loss to around one by the twentieth iteration. In the final iteration, the minimal training loss is achieved by RMSPROP, followed by SGDM and ADAM. Consequently, the testing accuracies of SqueezeNet model optimized by these optimizers also follow the same order (Table 5).

Fig. 15 shows training accuracy evolution of the optimizers for SqueezeNet model. As discussed previously, the training accuracy is the percentage estimation accuracy of a model in a particular training iteration. Training accuracy curves show how well the optimized deep learning models learn from their past accuracies and try to improve them further in the forthcoming iterations. The RMSPROP optimizer clearly achieves the highest accuracy at the end of the training iterations, followed by SGDM and ADAM.

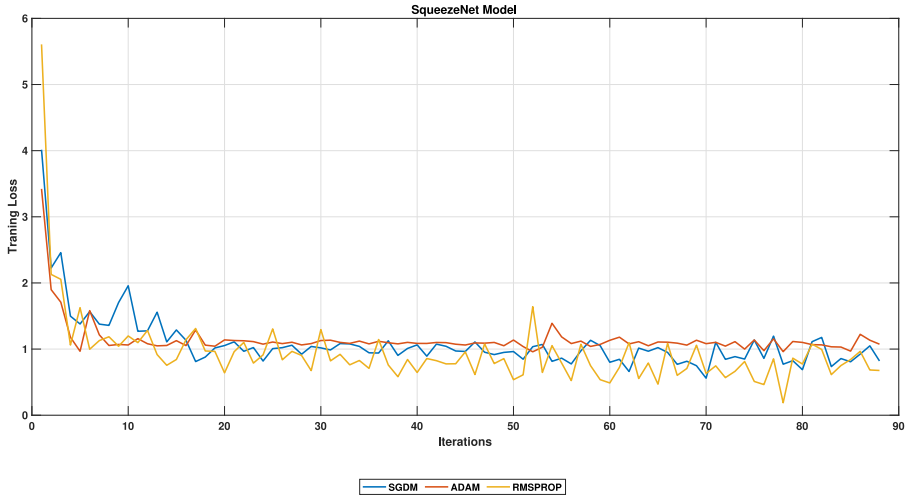


Fig. 14. Performance of training loss of optimizers in SqueezeNet for acoustic classification.

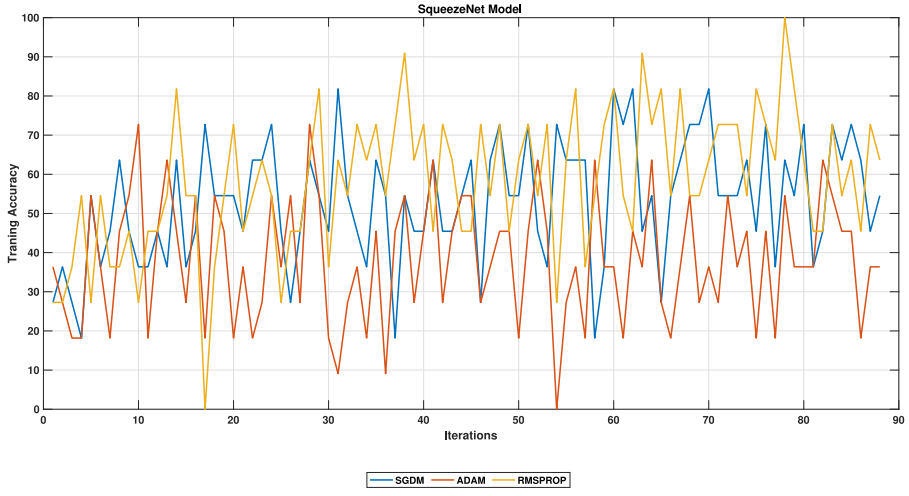


Fig. 15. Performance of training accuracy of optimizers in SqueezeNet for acoustic classification.

These accuracies are of the order of 65 %, 55 % and 35 % respectively, closely resembling the respective testing accuracies of these model-optimizer combinations.

3.4.2. Optimizers in GoogLeNet model

Fig. 16 shows that the both ADAM and RMSPROP were able to limit the GoogLeNet estimation error summation a little above one for most of the iterations. In the final step, the RMSPROP optimized GoogLeNet appeared to have gained a lower training loss than its ADAM counterpart. This could have contributed towards better testing accuracy achieved by RMSPROP GoogLeNet eventually. The SGDM optimizer was unable to successfully control the GoogLeNet training loss after the first few iterations.

Fig. 17 shows that ADAM and RMSPROP achieve minor improvements in GoogLeNet training accuracy over successive iterations. SGDM is unable to enhance GoogLeNet training accuracy. In the final step, highest training accuracy is achieved by the RMSPROP (around 45 %) followed by ADAM (around 37 %) and SGDM (around 27 %). The testing accuracies of these model optimizers resemble their final training accuracies respectively. The below average training accuracies of GoogLeNet models indicate major estimation errors made by these models. The minor training losses of ADAM and RMSPROP optimized GoogLeNet models indicate that these major estimation errors were made over a minor part of the training dataset. The SGDM optimized GoogLeNet model made major estimation errors (corresponding to its low training accuracy) over a large part of the dataset (corresponding to its high training loss) and hence, was unable to successfully classify the given dataset.

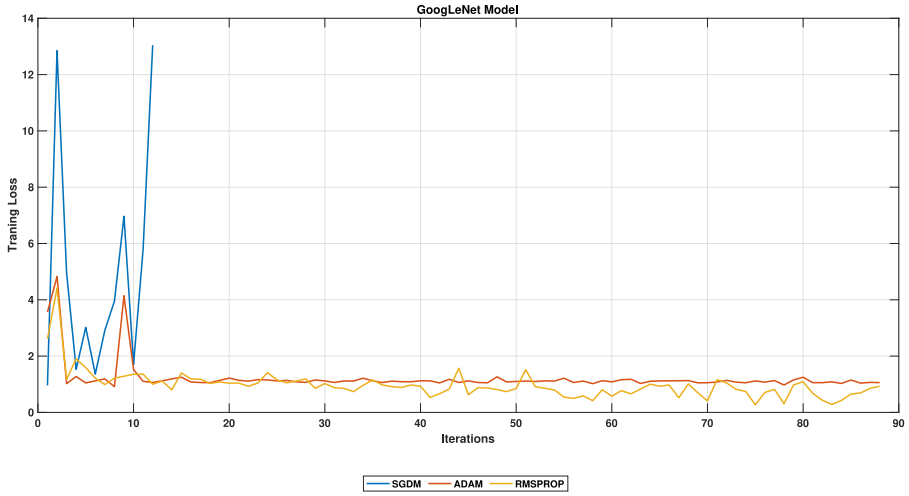


Fig. 16. Performance of training loss of optimizers in GoogLeNet for acoustic classification.

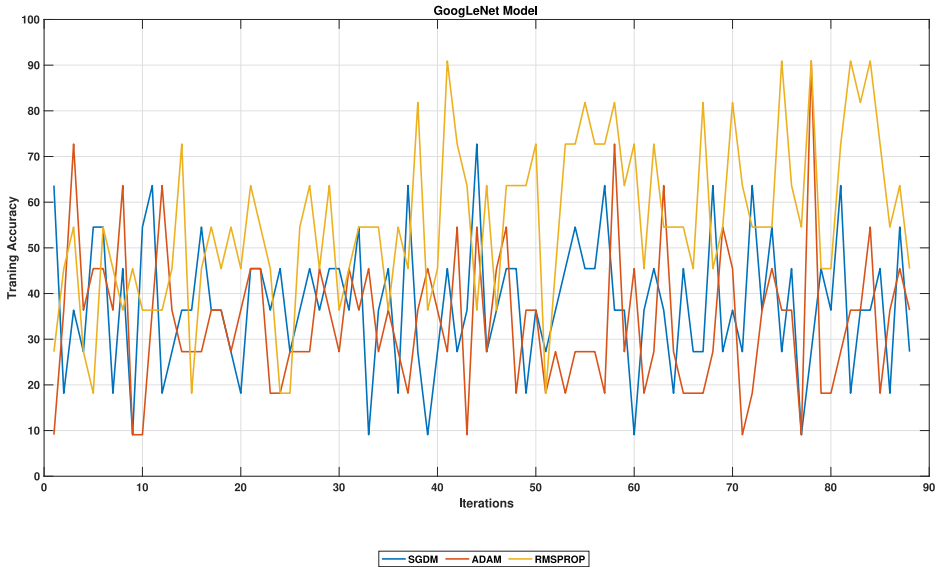


Fig. 17. Performance of training accuracy of optimizers in GoogLeNet for acoustic classification.

3.4.3. Optimizers in ResNet50 model

Fig. 18 shows the training loss progression of the ResNet50 models optimized by SGDM, ADAM and RMSPROP. All three optimizers successfully minimize ResNet50 estimation errors over successive iterations. SGDM and RMSPROP are able to limit training error to below one and below 0.5 respectively in the last iteration. ADAM optimized ResNet50 attains total estimation error of little above one in the final step.

Fig. 19 depicts the training accuracy growth of these model-optimizers. ADAM optimizer is unable to improve training accuracy of ResNet50, whereas SGDM could enhance it marginally. The RSMPROP registered an impressive hike in training accuracy, from about 45 % in the first step to more than 80 % in the final iteration. However, Table 5) indicates the lowest testing accuracy of RMSPROP optimized ResNet50 model as compared to its SGDM and ADAM optimized siblings. This indicates considerable overfitting of ResNet50 over the training dataset by RMSPROP. The higher training accuracies of SGDM-ResNet50 over SGDM-SqueezeNet, SGDM-GoogLeNet and that of RMSPROP-ResNet50 over RMSPROP-SqueezeNet and RMSPROP-GoogleLeNet indicates smaller estimation errors of the former configurations over the latter. Similarly, the lower training losses of ADAM and RMSPROP optimized ResNet50 as compared to the losses of their corresponding SqueezeNet and GoogLeNet optimized models indicate these errors to be present over a smaller part of the training dataset in case of the former models as compared to the latter.

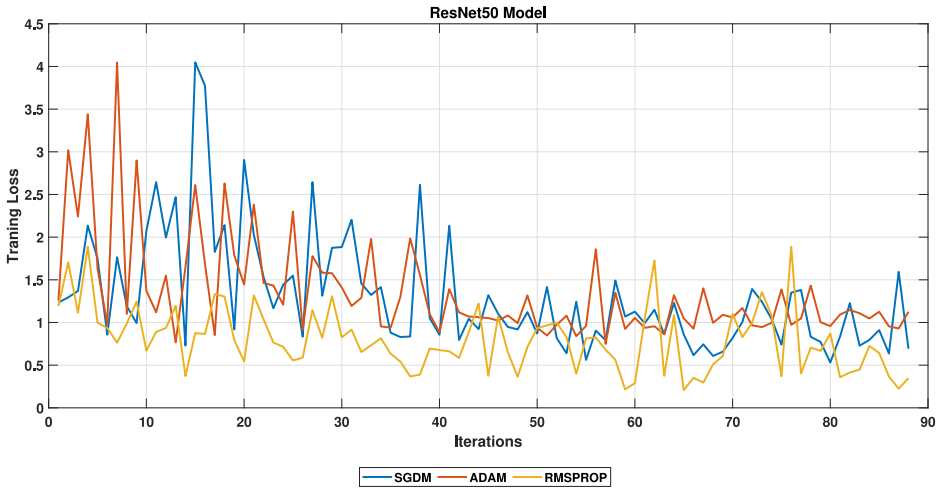


Fig. 18. Performance of training loss of optimizers in ResNet50 for acoustic classification.

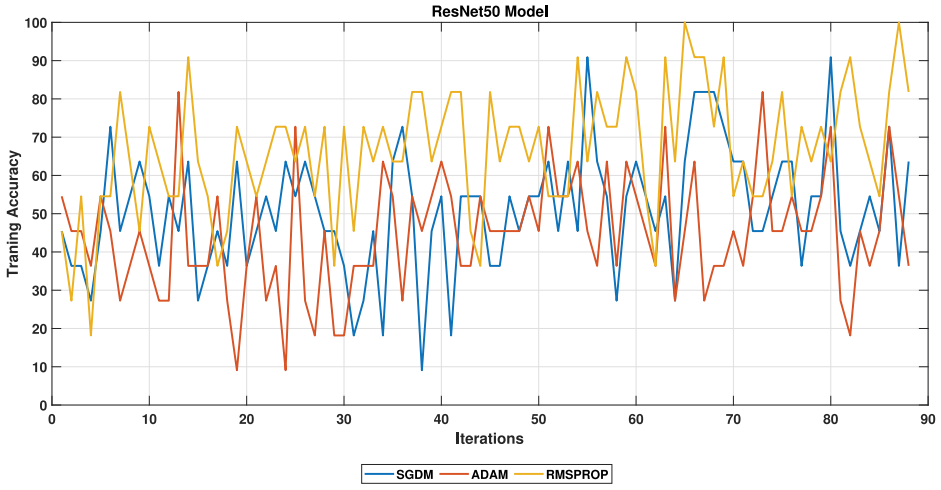


Fig. 19. Performance of training accuracy of optimizers in ResNet50 for acoustic classification.

3.4.4. Optimizers in InceptionV3 model

Fig. 20 depicts the training losses of the InceptionV3 model under various optimizers considered in the current work. The summation of estimation errors of this model is limited to one before the twentieth iteration by all optimizers. In the final step, the RMSPROP optimized InceptionV3 achieves the least training loss followed by its SGDM and ADAM counterparts. Table 5 shows that the testing accuracies of these models are attained in the same order.

Fig. 21 depicts the training accuracy evolution of these model-optimizers. The RMSPROP makes the most remarkable enhancement in accuracy over successive iterations, from less than 20 % in the first step to more than 80 % in the final iteration. The SGDM does not register any such progressive improvement and achieves the final accuracy of about 45 %. The ADAM optimizer finishes even lower (below 20 % accuracy). These final training accuracies resemble the respective testing accuracies of the said model-optimizers. Unlike for ResNet50, the RSMPROP does not overfit InceptionV3 model over the training dataset and achieves the best multi class classification testing accuracy of 87.50 %. Its low training loss coupled with high training accuracy indicate its smaller estimation errors over smaller part of the training dataset.

4. Conclusions, limitations, contributions and future scope

This paper presented an acoustic emission-texture extraction-transfer learning based carbide tool insert wear condition monitoring solution for CNC turning of mild steel rods. The mild steel specimens were machined through 80 experiments involving varying cutting speeds with constant feed rates and depths of cut. Three separate insert edges were used to consecutively machine the raw material through varying numbers of experimental runs involving progressively increasing cutting speeds. This experimental design

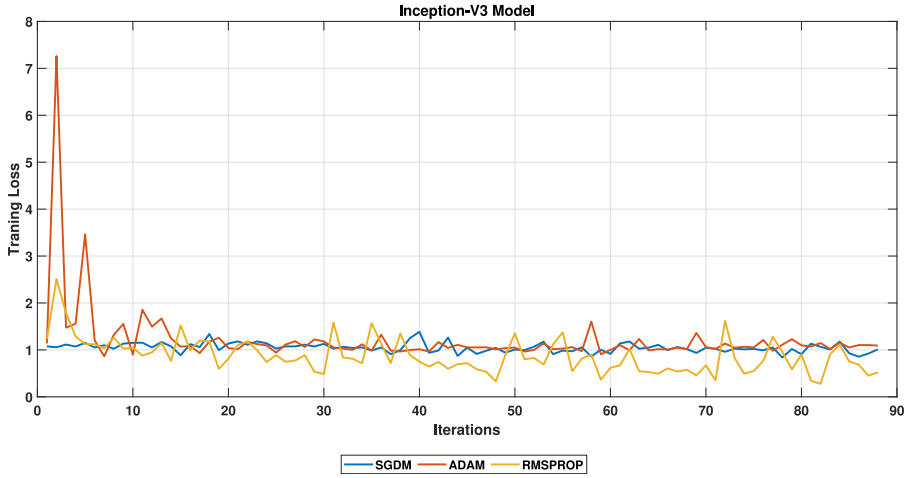


Fig. 20. Performance of training loss of optimizers in Inception-V3 for acoustic classification.

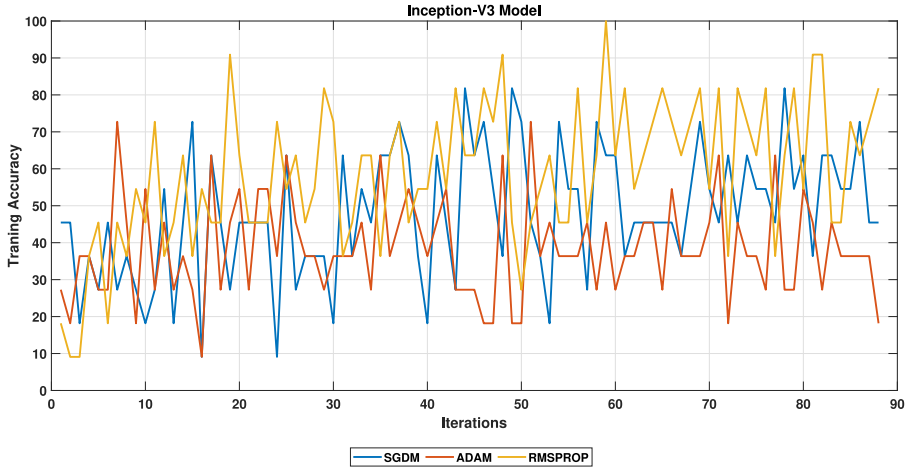


Fig. 21. Performance of training accuracy of optimizers in Inception-V3 for acoustic classification.

produced varying extents of tool wear on the insert edges at the end of their respective numbers of machining runs. Hence, the wear generated on these insert edges were labelled as low, medium and high. Acoustic emissions were recorded during all experimental runs using homogeneous smart phone microphone sensors. The noise recording of each experiment was labelled as either low, medium or high wear class. This labelling depended on the extent of wear expected to be generated on the tool inserts corresponding to the machining condition (cutting speed level) prevalent at that particular experimental run. In the next step, spectrogram images were obtained from all acoustic files using MFCC and sample rate filter techniques in Matlab. These spectrogram images were applied to a transfer learning architecture for three-class classification of tool wear. Four deep learning models viz. SqueezeNet, ResNet50, InceptionV3 and GoogLeNet were employed with three optimizers (RMSPROP, ADAM and SGDM) each for improved classification results. The performances of all model-optimizers were evaluated on the basis of testing accuracy, precision, recall, F1 scores, and the evolution of training loss/accuracy over successive training epochs.

Primary results indicate that the InceptionV3-RMSPROP obtained the highest testing accuracy of 87.50%, followed by the SqueezeNet-RMSPROP and ResNet50-SGDM with 75.00% and 62.50% testing accuracies respectively. However, SqueezeNet-RMSPROP was found to be more desirable for tool wear classification from a practical machining quality and safety point of view, owing to its greater recall value of high tool wear class over that of the InceptionV3-RMSPROP and ResNet50-SGDM model-optimizers. The limitations of this study include the following:

1. The first limitation in this study was imposed by the relatively lesser number of samples of the middle-level tool wear class dataset used in the present study, which on one hand did not allow sufficient classification accuracies for most of the model-optimizers, but on the other hand revealed that the RMSPROP optimized models could perform well regardless, highlighting the superior ability of RMSPROP optimized models to handle not only smaller but unbalanced datasets as well.

2. The second important limitation lies in the fact that more deep learning models and/or optimizers need to be explored in search of even better classification accuracies for the given dataset, which is smaller in size and is unbalanced as well.

In a nutshell, this study demonstrates the following contributions and scope for further investigations:

1. Acoustic data corresponding to tool wear modes can be easily collected and converted to spectrograms for image based classification, to ultimately build a robust tool condition monitoring system that takes care of machining quality and safety as well.
2. Transfer learning enables application of deep learning models and optimizers to smaller (and even unbalanced) datasets with reasonably good accuracy.
3. It is important to take into account the respective precision and recall values (and not just testing accuracies) of deep learning model-optimizer combinations to take care of the practical machining aspects.
4. **Future scope:** While classifying smaller datasets, the deep learning models with the capability of collapsing/squeezing larger layers/ greater parameters into smaller layers/ lesser parameters must be employed in conjunction with adaptive rate equipped and stable learning optimizers (such as RMSPROP) for better results in future research endeavours.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

[1] Kane P, Andhare A. Critical evaluation and comparison of psychoacoustics, acoustics and vibration features for gear fault correlation and classification. Measurement 2020;107495.
[2] Delvecchio S, Bonfiglio P, Pompoli F. Vibro-acoustic condition monitoring of internal combustion engines: A critical review of existing techniques. Mech Syst Signal Process 2018;99:661–83. http://dx.doi.org/10.1016/j.ymssp.2017.06.033, URL http://www.sciencedirect.com/science/article/pii/S088832701730345X.
[3] Kene AP, Choudhury SK. Analytical modeling of tool health monitoring system using multiple sensor data fusion approach in hard machining. Measurement 2019;145:118–29.
[4] Bhuiyan M, Choudhury I. 13.22—Review of sensor applications in tool condition monitoring in machining. Comprehens Mater Process 2014;13:539–69.
[5] Alatorre D, Rabani A, Axinte D, Branson DT. Closed loop force control of in-situ machining robots using audible sound features. Mech Syst Signal Process 2020;136:106517.
[6] Lee WJ, Wu H, Yun H, Kim H, Jun MB, Sutheralnd JW. Predictive maintenance of machine tool systems using artificial intelligence techniques applied to machine condition data. Procedia CIRP 2019;80:506–11.
[7] Galarza-Urigoitia N, Rubio-García B, Gascón-Álvarez J, Aznar-Lapuente G, Olite-Biurrun J, López-Germán A, et al. Predictive maintenance of wind turbine low-speed shafts based on an autonomous ultrasonic system. Eng Fail Anal 2019;103:481–504.
[8] Khan MA, Shahid MA, Ahmed SA, Khan SZ, Khan KA, Ali SA, et al. Gear misalignment diagnosis using statistical features of vibration and airborne sound spectrums. Measurement 2019;145:419–35.
[9] Becht P, Deckers E, Claeys C, Pluymers B, Desmet W. Loose bolt detection in a complex assembly using a vibro-acoustic sensor array. Mech Syst Signal Process 2019;130:433–51. http://dx.doi.org/10.1016/j.ymssp.2019.05.019, URL http://www.sciencedirect.com/science/article/pii/S0888327019303280.
[10] Schnabel S, Marklund P, Larsson R, Golling S. The detection of plastic deformation in rolling element bearings by acoustic emission. Tribol Int 2017;110:209–15. http://dx.doi.org/10.1016/j.triboint.2017.02.021, URL http://www.sciencedirect.com/science/article/pii/S0301679X17300816.
[11] del Campo SM, Schnabel S, Sandin F, Marklund P. Detection of particle contaminants in rolling element bearings with unsupervised acoustic emission feature learning. Tribol Int 2019;132:30–8. http://dx.doi.org/10.1016/j.triboint.2018.12.007, URL http://www.sciencedirect.com/science/article/pii/S0301679X18305772.
[12] Fuentes R, Dwyer-Joyce R, Marshall M, Wheals J, Cross E. Detection of sub-surface damage in wind turbine bearings using acoustic emissions and probabilistic modelling. Renew Energy 2020;147:776–97. http://dx.doi.org/10.1016/j.renene.2019.08.019, URL http://www.sciencedirect.com/science/article/pii/S0960148119312066.
[13] Dykas B, Harris J. Acoustic emission characteristics of a single cylinder diesel generator at various loads and with a failing injector. Mech Syst Signal Process 2017;93:397–414. http://dx.doi.org/10.1016/j.ymssp.2017.01.049, URL http://www.sciencedirect.com/science/article/pii/S0888327017300547.
[14] Dias EA, Pereira FB, Filho SLMR, Brandão LC. Monitoring of through-feed centreless grinding processes with acoustic emission signals. Measurement 2016;94:71–9. http://dx.doi.org/10.1016/j.measurement.2016.07.075, URL http://www.sciencedirect.com/science/article/pii/S0263224116304389.
[15] Becht P, Deckers E, Claeys C, Pluymers B, Desmet W. Selection of small sensor arrays for localization of damage in complex assemblies using vibro-acoustic signals. In: Wahab MA, editor. Proceedings of the 13th international conference on damage assessment of structures. Singapore: Springer Singapore; 2020, p. 263–82.
[16] Twardowski P, Tabaszewski M, Wiciak – Pikuła M, Felusiak-Czyryca A. Identification of tool wear using acoustic emission signal and machine learning methods. Precis Eng 2021;72:738–44. http://dx.doi.org/10.1016/j.precisioneng.2021.07.019, URL https://www.sciencedirect.com/science/article/pii/S0141635921001884.
[17] Straka L, Čorný I. Identification of geometric errors of circular profiles at WEDM caused by the wire tool electrode vibrations and their reduction with support of acoustic emission method. Eng Fail Anal 2022;134:106040. http://dx.doi.org/10.1016/j.engfailanal.2022.106040, URL https://www.sciencedirect.com/science/article/pii/S1350630722000140.

[18] Cooper C, Zhang J, Gao RX, Wang P, Ragai I. Anomaly detection in milling tools using acoustic signals and generative adversarial networks. Procedia Manuf 2020;48:372–8. http://dx.doi.org/10.1016/j.promfg.2020.05.059, 48th SME North American Manufacturing Research Conference, NAMRC 48, URL https://www.sciencedirect.com/science/article/pii/S2351978920315110.

[19] Cooper C, Wang P, Zhang J, Gao RX, Roney T, Ragai I, et al. Convolutional neural network-based tool condition monitoring in vertical milling operations using acoustic signals. Procedia Manuf 2020;49:105–11. http://dx.doi.org/10.1016/j.promfg.2020.07.004, Proceedings of the 8th International Conference on Through-Life Engineering Services – TESConf 2019, URL https://www.sciencedirect.com/science/article/pii/S2351978920316565.

[20] Wan L, Zhang X, Zhou Q, Wen D, Ran X. Acoustic emission identification of wheel wear states in engineering ceramic grinding based on parameter-adaptive VMD. Ceram Int 2022. http://dx.doi.org/10.1016/j.ceramint.2022.12.238, URL https://www.sciencedirect.com/science/article/pii/S0272884222046697.

[21] Xu X, Yang Z, Liu Q, Yan S, Ding H. Condition monitoring and mechanism analysis of belt wear in robotic grinding of TC4 workpiece using acoustic emissions. Mech Syst Signal Process 2023;188:109979. http://dx.doi.org/10.1016/j.ymssp.2022.109979, URL https://www.sciencedirect.com/science/article/pii/S0888327022010470.

[22] Revill P, Clarke A, Pullin R, Dennis G. Acoustic emission monitoring of wear in aerospace self-lubricating bearing liner materials. Wear 2021;486–487:204102. http://dx.doi.org/10.1016/j.wear.2021.204102, URL https://www.sciencedirect.com/science/article/pii/S0043164821004865.

[23] Shen C-H. Acoustic emission based grinding wheel wear monitoring: Signal processing and feature extraction. Appl Acoust 2022;196:108863. http://dx.doi.org/10.1016/j.apacoust.2022.108863, URL https://www.sciencedirect.com/science/article/pii/S0003682X22002377.

[24] Deshpande P, Pandiyan V, Meylan B, Wasmer K. Acoustic emission and machine learning based classification of wear generated using a pin-on-disc tribometer equipped with a digital holographic microscope. Wear 2021;476:203622. http://dx.doi.org/10.1016/j.wear.2021.203622, 23rd International Conference on Wear of Materials, URL https://www.sciencedirect.com/science/article/pii/S0043164821000119.

[25] Unterberg M, Voigts H, Weiser IF, Feuerhack A, Trauth D, Bergs T. Wear monitoring in fine blanking processes using feature based analysis of acoustic emission signals. Procedia CIRP 2021;104:164–9. http://dx.doi.org/10.1016/j.procir.2021.11.028, 54th CIRP CMS 2021 - Towards Digitalized Manufacturing 4.0, URL https://www.sciencedirect.com/science/article/pii/S2212827121009264.

[26] Kuntoğlu M, Sağlam H. ANOVA and fuzzy rule based evaluation and estimation of flank wear, temperature and acoustic emission in turning. CIRP J Manuf Sci Technol 2021;35:589–603. http://dx.doi.org/10.1016/j.cirpj.2021.07.011, URL https://www.sciencedirect.com/science/article/pii/S1755581721001267.

[27] König F, Marheineke J, Jacobs G, Sous C, Zuo MJ, Tian Z. Data-driven wear monitoring for sliding bearings using acoustic emission signals and long short-term memory neural networks. Wear 2021;476:203616. http://dx.doi.org/10.1016/j.wear.2021.203616, 23rd International Conference on Wear of Materials, URL https://www.sciencedirect.com/science/article/pii/S0043164821000053.

[28] Ahmed YS, Arif A, Veldhuis S. Application of the wavelet transform to acoustic emission signals for built-up edge monitoring in stainless steel machining. Measurement 2020;107478.

[29] Solke NS, Shah P, Sekhar R, Singh T. Machine learning-based predictive modeling and control of lean manufacturing in automotive parts manufacturing industry. Global J Flexible Syst Manag 2022;23(1):89–112.

[30] Sekhar R, Solke N, Shah P. Lean manufacturing soft sensors for automotive industries. Appl Syst Innov 2023;6(1). URL https://www.mdpi.com/2571-5577/6/1/22.

[31] Sekhar R, Shah P. Predictive modeling of a flexible robotic arm using cohort intelligence socio-inspired optimization. In: 2020 1st International conference on information technology, advanced mechanical and electrical engineering. IEEE; 2020, p. 193–8.

[32] Hu Z-X, Wang Y, Ge M-F, Liu J. Data-driven fault diagnosis method based on compressed sensing and improved multiscale network. IEEE Trans Ind Electron 2020;67(4):3216–25. http://dx.doi.org/10.1109/TIE.2019.2912763.

[33] Sekhar R, Singh T, Shah P. Machine learning based predictive modeling and control of surface roughness generation while machining micro boron carbide and carbon nanotube particle reinforced al-mg matrix composites. Particul Sci Technol 2021;1–18.

[34] Liu J, Hu Y, Wang Y, Wu B, Fan J, Hu Z. An integrated multi-sensor fusion-based deep feature learning approach for rotating machinery diagnosis. Meas Sci Technol 2018;29(5):055103. http://dx.doi.org/10.1088/1361-6501/aaaca6.

[35] Purohit K, Srivastava S, Nookala V, Joshi V, Shah P, Sekhar R, et al. Soft sensors for state of charge, state of energy and power loss in formula student electric vehicle. Appl Syst Innov 2021;4(4):78.

[36] Sekhar R, Shah P, Panchal S, Fowler M, Fraser R. Distance to empty soft sensor for ford escape electric vehicle. Results Control Optim 2022;9:100168. http://dx.doi.org/10.1016/j.rico.2022.100168, URL https://www.sciencedirect.com/science/article/pii/S2666720722000406.

[37] Shah P, Sharma D, Sekhar R. Analysis of research trends in fractional controller using latent Dirichlet allocation. Eng Lett 2021;29(1).

[38] Sharma D, Kumar B, Chand S, Shah RR. A trend analysis of significant topics over time in machine learning research. SN Comput Sci 2021;2(6):1–13.

[39] Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the Inception architecture for computer vision. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016, p. 2818–26. http://dx.doi.org/10.1109/CVPR.2016.308.

[40] Szegedy C, Ioffe S, Vanhoucke V, Alemi AA. Inception-v4, Inception-ResNet and the impact of residual connections on learning. In: Thirty-first AAAI conference on artificial intelligence. 2017, URL https://arxiv.org/abs/1602.07261.

[41] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. 2014, arXiv preprint arXiv:1409.1556, URL https://arxiv.org/abs/1409.1556.

[42] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2015, p. 1–9. http://dx.doi.org/10.1109/CVPR.2015.7298594.

[43] Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2014, p. 580–7, URL http://arxiv.org/abs/1311.2524.

[44] Wang N, Yeung DY. Learning a deep compact image representation for visual tracking. Adv Neural Inf Process Syst 2013. URL https://proceedings.neurips.cc/paper/2013/file/dc6a6489640ca02b0d42dabeb8e46bb7-Paper.pdf.

[45] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2015, p. 3431–40. http://dx.doi.org/10.1109/TPAMI.2016.2572683.

[46] Dong C, Loy CC, He K, Tang X. Learning a deep convolutional network for image super-resolution. In: European conference on computer vision. Springer; 2014, p. 184–99. http://dx.doi.org/10.1007/978-3-319-10593-2_13.

[47] Karpathy A, Toderici G, Shetty S, Leung T, Sukthankar R, Fei-Fei L. Large-scale video classification with convolutional neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2014, p. 1725–32. http://dx.doi.org/10.1109/CVPR.2014.223.

[48] Toshev A, Szegedy C. Human pose estimation via deep neural networks'. In: CVPR(Columbus, Ohio, 2014). 2014, p. 1653–60. http://dx.doi.org/10.1109/CVPR.2014.214.

[49] Erhan D, Szegedy C, Toshev A, Anguelov D. Scalable object detection using deep neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2014, p. 2147–54, URL http://arxiv.org/abs/1312.2249.

[50] Sandler M, Howard A, Zhu M, Zhmoginov A, Chen L-C. MobileNetV2: Inverted residuals and linear bottlenecks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2019, p. 4510–20. http://dx.doi.org/10.1109/CVPR.2018.00474.

[51] Zhang X, Zhou X, Lin M, Sun J. ShuffleNet: An extremely efficient convolutional neural network for mobile devices. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2018, p. 6848–56. http://dx.doi.org/10.1109/CVPR.2018.00716.

[52] Changpinyo S, Sandler M, Zhmoginov A. The power of sparsity in convolutional neural networks. 2017, arXiv preprint arXiv:1702.06257, URL https://arxiv.org/abs/1702.06257.

[53] Hassibi B, Stork DG. Second order derivatives for network pruning: optimal brain surgeon, vol. 5, Morgan Kaufmann; 1993, URL https://proceedings.neurips.cc/paper/1992/file/303ed4c69846ab36c2904d3ba8573050-Paper.pdf.
[54] LeCun Y, Denker JS, Solla SA. Optimal brain damage. In: Advances in neural information processing systems. 1990, p. 598–605, URL http://yann.lecun.com/exdb/publis/pdf/lecun-90b.pdf.
[55] Han S, Pool J, Tran J, Dally WJ. Learning both weights and connections for efficient neural networks. 2015, arXiv preprint arXiv:1506.02626, URL http://arxiv.org/abs/1506.02626.
[56] Han S, Pool J, Narang S, Mao H, Tang S, Elsen E, et al. DSD: regularizing deep neural networks with Dense-Sparse-Dense training flow. 2016, URL http://arxiv.org/abs/1607.04381.
[57] Guo Y, Yao A, Chen Y. Dynamic network surgery for efficient DNNs. 2016, arXiv preprint arXiv:1608.04493 URL http://arxiv.org/abs/1608.04493.
[58] Li H, Kadav A, Durdanovic I, Samet H, Graf HP. Pruning filters for efficient convnets. 2016, arXiv preprint arXiv:1608.08710, URL https://arxiv.org/abs/1608.08710.
[59] Ahmed K, Torresani L. Connectivity learning in multi-branch networks. 2017, arXiv preprint arXiv:1709.09582 http://arxiv.org/abs/1709.09582.
[60] Veniat T, Denoyer L. Learning time/memory-efficient deep architectures with budgeted super networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2018, p. 3492–500. http://dx.doi.org/10.1109/CVPR.2018.00368.
[61] Bergstra J, Bengio Y. Random search for hyper-parameter optimization.. J Mach Learn Res 2012;13(2). URL https://www.jmlr.org/papers/volume13/bergstra12a/bergstra12a.pdf.
[62] Snoek J, Larochelle H, Adams RP. Practical bayesian optimization of machine learning algorithms. Adv Neural Inf Process Syst 2012;25.
[63] Snoek J, Rippel O, Swersky K, Kiros R, Satish N, Sundaram N, et al. Scalable bayesian optimization using deep neural networks. In: International conference on machine learning. PMLR; 2015, p. 2171–80, URL https://arxiv.org/abs/1502.05700.
[64] Shah P, Sekhar R, Kulkarni AJ, Siarry P. Metaheuristic algorithms in industry 4.0. CRC Press; 2021.
[65] Zoph B, Vasudevan V, Shlens J, Le QV. Learning transferable architectures for scalable image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2018, p. 8697–710. http://dx.doi.org/10.1109/CVPR.2018.00907.
[66] Xie L, Yuille A. Genetic CNN. In: Proceedings of the IEEE international conference on computer vision. 2017, p. 1379–88, URL https://arxiv.org/abs/1703.01513.
[67] Real E, Moore S, Selle A, Saxena S, Suematsu YL, Tan J, et al. Large-scale evolution of image classifiers. In: International conference on machine learning. PMLR; 2017, p. 2902–11, URL https://arxiv.org/abs/1703.01041.
[68] Zoph B, Le QV. Neural architecture search with reinforcement learning. 2016, arXiv preprint arXiv:1611.01578 URL https://arxiv.org/abs/1611.01578.
[69] Sekhar R, Sharma D, Shah P. Intelligent classification of tig welding defects: A transfer learning approach. 2022, URL https://www.frontiersin.org/articles/10.3389/fmech.2022.824038/abstract,
[70] Hussain M, Bird JJ, Faria DR. A study on CNN transfer learning for image classification. In: UK workshop on computational intelligence. Springer; 2018, p. 191–202. http://dx.doi.org/10.1007/978-3-319-97982-3_16.
[71] Gao Y, Mosalam KM. Deep transfer learning for image-based structural damage recognition. Comput-Aided Civ Infrastruct Eng 2018;33(9):748–68. http://dx.doi.org/10.1111/mice.12363.
[72] Larsen-Freeman D. Transfer of learning transformed. Lang Learn 2013;63:107–29. http://dx.doi.org/10.1111/j.1467-9922.2012.00740.x.
[73] Shaha M, Pawar M. Transfer learning for image classification. In: 2018 Second international conference on electronics, communication and aerospace technology. 2018, p. 656–60. http://dx.doi.org/10.1109/ICECA.2018.8474802.
[74] Ajmi C, Zapata J, Elferchichi S, Zaafouri A, Laabidi K. Deep learning technology for weld defects classification based on transfer learning and activation features. Adv Mater Sci Eng 2020;2020. http://dx.doi.org/10.1155/2020/1574350.
[75] Zhu Y, Chen Y, Lu Z, Pan SJ, Xue G-R, Yu Y, et al. Heterogeneous transfer learning for image classification. In: Twenty-fifth AAAI conference on artificial intelligence. 2011, URL https://www.aaai.org/ocs/index.php/AAAI/AAAI11/paper/viewFile/3671/4073.
[76] Li W, Huang R, Li J, Liao Y, Chen Z, He G, et al. A perspective survey on deep transfer learning for fault diagnosis in industrial scenarios: Theories, applications and challenges. Mech Syst Signal Process 2022;167:108487. http://dx.doi.org/10.1016/j.ymssp.2021.108487.
[77] Lockner Y, Hopmann C, Zhao W. Transfer learning with artificial neural networks between injection molding processes and different polymer materials. J Manuf Process 2022;73:395–408. http://dx.doi.org/10.1016/j.jmapro.2021.11.014.
[78] Yang X, Zhang Y, Lv W, Wang D. Image recognition of wind turbine blade damage based on a deep learning model with transfer learning and an ensemble learning classifier. Renew Energy 2021;163:386–97. http://dx.doi.org/10.1016/j.renene.2020.08.125.
[79] Yin H, Ou Z, Fu J, Cai Y, Chen S, Meng A. A novel transfer learning approach for wind power prediction based on a serio-parallel deep learning architecture. Energy 2021;234:121271. http://dx.doi.org/10.1016/j.energy.2021.121271.
[80] Bhatt A, Ganatra A, Kotecha K. COVID-19 pulmonary consolidations detection in chest X-ray using progressive resizing and transfer learning techniques. Heliyon 2021. http://dx.doi.org/10.1016/j.heliyon.2021.e07211.
[81] Tanveer MH, Zhu H, Ahmed W, Thomas A, Imran BM, et al. Mel-spectrogram and deep cnn based representation learning from bio-sonar implementation on uavs. In: 2021 International conference on computer, control and robotics. IEEE; 2021, p. 220–4.
[82] Sharma K, Reddy S. Spectrogram analysis and text conversion of sound signal for query generation to give input to audio input device. In: Artificial intelligence and speech technology: third international conference, AIST 2021, Delhi, India, November 12–13, 2021, revised selected papers. Springer; 2022, p. 171–81.
[83] Zhivomirov H. Sound analysis with matlab implementation. Sound Anal Matlab Implement 2010.
[84] Sekhar R, Shah P, Panchal S, Fowler M, Fraser R. Distance to empty soft sensor for ford escape electric vehicle. Results Control Optim 2022;9:100168.
[85] Koonce B, Koonce B. SqueezeNet. Convolut Neural Netw Swift Tensorflow: Image Recogn Dataset Categoriz 2021;73–85.
[86] Iandola FN, Han S, Moskewicz MW, Ashraf K, Dally WJ, Keutzer K. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size. 2016, arXiv preprint arXiv:1602.07360.
[87] Ucar F, Korkmaz D. Covidiagnosis-net: Deep Bayes-SqueezeNet based diagnosis of the coronavirus disease 2019 (COVID-19) from X-ray images. Med Hypotheses 2020;140:109761.
[88] Alzubaidi L, Zhang J, Humaidi AJ, Al-Dujaili A, Duan Y, Al-Shamma O, et al. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. J Big Data 2021;8:1–74.
[89] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016, p. 770–8, URL http://arxiv.org/abs/1512.03385.
[90] Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, et al. MobileNets: Efficient convolutional neural networks for mobile vision applications. 2017, arXiv preprint arXiv:1704.04861, URL http://arxiv.org/abs/1704.04861.
[91] Tieleman T, Hinton G. Divide the gradient by a running average of its recent magnitude. coursera: Neural networks for machine learning. Technical report, 2017.
[92] Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. ImageNet large scale visual recognition challenge. Int J Comput Vision 2015;115(3):211–52. http://dx.doi.org/10.1007/s11263-015-0816-y.
[93] He K, Zhang X, Ren S, Sun J. Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification. In: Proceedings of the IEEE international conference on computer vision. 2015, p. 1026–34, URL http://arxiv.org/abs/1502.01852.

[94] Robbins H, Monro S. A stochastic approximation method. Annals Math Statist 1951;400–7. http://dx.doi.org/10.1214/aoms/1177729586.
[95] Kingma DP, Ba J. Adam: A method for stochastic optimization. 2014, arXiv preprint arXiv:1412.6980, URL https://arxiv.org/abs/1412.6980.
[96] Sekhar R, Sharma D, Shah P. Intelligent classification of tungsten inert gas welding defects: A transfer learning approach. Front Mech Eng 2022;8:824038.
[97] Bacioiu D, Melton G, Papaelias M, Shaw R. Automated defect classification of SS304 TIG welding process using visible spectrum camera and machine learning. NDT & E Int 2019;107:102139. http://dx.doi.org/10.1016/j.ndteint.2019.102139.