3.3.1: Number of research papers published per teacher in the Journals as notified on UGC CARE list during the last five years

3.3.1.1. Number of research papers in the Journals notified on UGC CARE list year wise during the last five years.

CALENDAR YEAR 2021

Title of paper	Name of the author/s	Department of the teacher	Name of journal	Calendar Year of publication	ISSN number	Link to the recognition in UGC enlistment of the Journal /Digital Object Identifier (doi) number		
						Link to website of the Journal	Link to article / paper / abstract of the article	Is it listed in UGC Care list
Evaluation of stress concentration factor for a plate with different position of the hole	Dr. Shailesh Patel	Department of Humanities and Sciences	International Journal of creative Research Thoughts	Apr-21	2320-2882	https://ijcrt.org/	https://ijcrt.org/viewfull.php?&p_id=IJCRT2104493	Yes
A Study on various Human Facial Feature Extraction Techniques in High Dimensional Spaces	Prof. Jaimin Jani	Department of Computer Engineering & Information Technology	Online International Conference on Recent Advances in Deep Learning (ICRADL- 2021)	Jan-21	ISSN Online: 2278-0181	https://www.ijert.org	https://www.ijert.org/a-study-on-various-human-facial-feature-extraction-techniques-in-high-	Yes

Title of paper	Name of the author/s	Department of the teacher	Name of journal	Calendar Year of publication	ISSN number	Link to the recognition in UGC enlistment of the Journal /Digital Object Identifier (doi) number		
						Link to website of the Journal	Link to article / paper / abstract of the article	Is it listed in UGC Care list
							dimensional-spaces	
A Study to Develop Human Face Recognition Using PCA, Neural Networks and Wavelet	Prof. Jaimin Jani	Department of Computer Engineering & Information Technology	Online International Conference on Recent Advances in Deep Learning (ICRADL- 2021)	Jan-21	ISSN: 2278-0181	https://www.ijert.org/	https://www.ijert.org/a-study-to-develop-human-face-recognition-using-pca-neural-networks-and-wavelet	Yes
Predictive Maintenance of Gas Turbine Using Prognosis Approach	Prof. Hima Soni	Department of Computer Engineering & Information Technology	International Research Journal of Engineering and Technology	Jan-21	2395-0056	https://www.irjet.net	https://www.irjet.net/archives/V7/i6/IRJET-V7I6878.pdf	Yes

IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

EVALUATION OF STRESS CONCENTRATION FACTOR FOR A PLATE WITH DIFFERENT POSTION OF THE HOLE

Prachi Patel [1],M.Y. Patil [2] Hiren R Prajapati[3], Shailesh T Patel[4]

[1]Research Scholar, Mechanical Engineering, at GEC Dahod, Gujarat,India

[2]Associate Professor, Mechanical Engineering Department,Government Engineering college Dahod, Gujarat, India

[3] Asst. Professor, Mechanical Engineering Department,Government Engineering college Modasa, Gujarat, India

[4] Asst. Professor, S.P.B.Patel Engineering college Mehsana, Gujarat, India

Abstract: The stress concentration factors are widely used to predict the maximum stress value above which the mechanical structure can be destroyed. Many chart data of those factors are available in literature but they are conditioned by the structure shape and the principal geometric dimensions. This paper compares, for thin plate with eccentric hole, the stress concentration factors values calculated by classical formulas given in ulterior studies and a numerical simulation using commercial software. The effects of the relative hole position with central hole in axial loading and compressive loading, varing the distance along the x axis, Displacement along the x-axis the in the plate and the in near the edge hole axis are examined.

***Index Terms* - Axial tension; finite element analysis; stress concentration factor;Hole equivalent (Von Mises) stress.**

I. INTRODUCTION

In industries many unexpected failure of machine components and equipments have occurred. These failures of machines components and equipments are due to either poor design or sudden changes that happen during working conditions. However, in so many cases, pre-existing geometrical irregularities are also a major causes for the failures. The geometric discontinuities such as keyways, notches, shoulders fillets, holes, various grooves (like U, V, square), threads etc. on different machine parts (i.e. on plates) are unavoidable due to their functional requirementThe plates with discontinuities like circular holes exist in all metal structures. Those areas represent dangerous zones because of the multiplication of the stresses values under the effect of the stress concentration phenomenon. These stress concentration zones are often areas of crack initiation. They can be dangerous if the loading conditions allow the brutal propagation of the cracks and than promote the rupture. The stresses concentration phenomenon is measured by a parameter called stress concentration factor (SCF), noted Kt. This factor is the ratio of the maximum elastic stress value by the nominal stress calculated in the discontinuity area. The values of this factor are calculated using analytical approaches based on the stress and deformation distributions evaluation around the discontinuity or by numerical models or also by experimental studies using the photoelasticity method. The results of these investigations are resumed in curves according to the structure geometry dimensions.

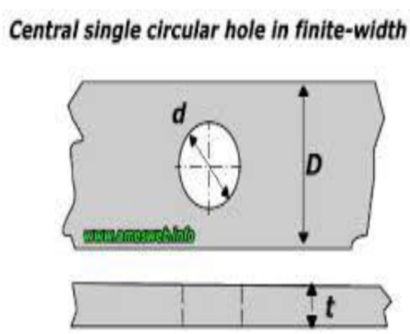
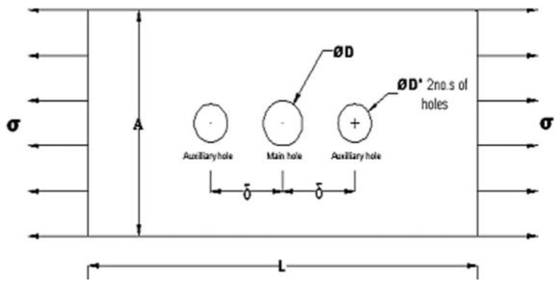


Fig : 1 Central hole with plate

Fig : 2 Rectangular plate with Central hole with Auxiliary hole[1]

II.STRESS CONCENTRATION FACTOR CALCULATION



The localization of high stress is known as stress concentration and is evaluated by a factor known as stress concentration factor (SCF). Common definition of stress concentration factor (SCF) is the ratio of the maximum stress (σ_{max}) to the nominal stress(σ_{nom}). Mathematically SCF can be written as,

$$k_t = \frac{\sigma_{max}}{\sigma_{nom}}$$

For plate with hole/holes stress concentration factor (SCF) is defined by two different way. Stress concentration factor is K_{tg}, for which the reference stress is based on the gross cross-sectional area, or K_{tn}, for which the reference stress is based on the net cross-sectional area [2].

$$K_{tg} = \frac{\sigma_{max}}{\sigma}$$

K_{tg} is the stress concentration factor based on gross stress, is the maximum stress, at the edge of the hole, σ is the stress on gross section far from the hole, and

$$K_{tn} = \frac{\sigma_{max}}{\sigma_n}$$

Where,K_{tn} is the stress concentration factor based on net (nominal) stress and σ_n is the net stress $\sigma/(1-d/H)$, with d the hole diameter and H the width of element.

The equations presented in this paragraph are widely used in the literature. However, it does not take into account all dimensions of the plate, as indicated in Fig. 1, like the length D, the thickness t and distance b. In the finite elements model, presented in the following section, all mentioned dimensions are specified and consequently the numerical simulations results are influenced.

Chart 1 shows stress concentration factors $\boldsymbol{K_{tg}}$ and $\boldsymbol{K_{tn}}$ for the tension of a finite-width thin element with a circular hole.

Chart 2 shows Stress concentration factors for the tension of a thin semi-infinite element with a circular hole near the edge [2].

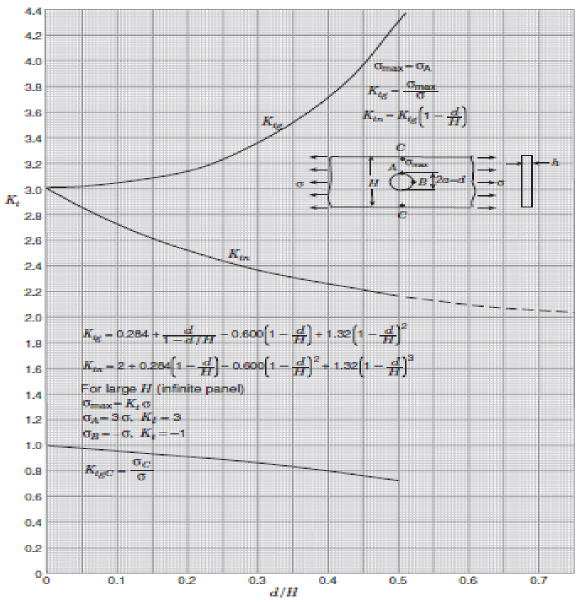


Chart 1 1 Stress concentration factors K_{tg} and K_{tn} for the tension of a finite-width thin element with a circular hole (Howland 1929–1930).(From Peterson's Handbook) [2]

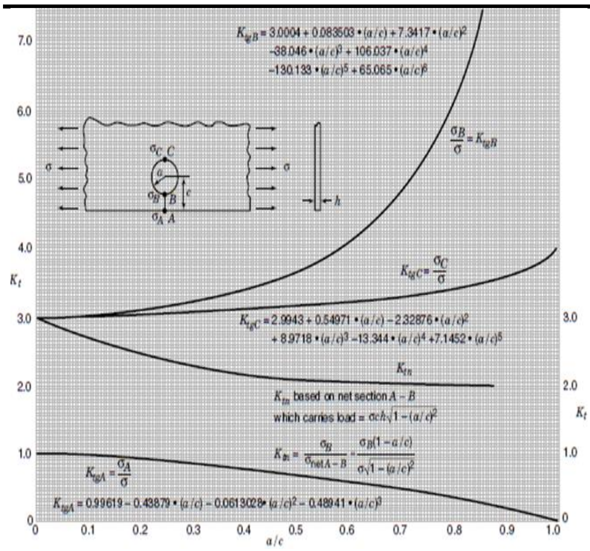


Chart 2 Stress concentration factors for the tension of a thin semi-infinite element with a circular hole near the edge (From Peterson's Handbook) [2]

III.LITERATUREREVIEW

The immense literature review has beendone on analysis of stress concentrationon a plate by considering hole position,materials for plate and applied load.

OZKANETAL.(DETERMINED THE STRESS CONCENTRATION FACTOR (KT) IN A RECTANGULAR PLATE WITH A HOLEUNDERTENSILE STRESS USING DIFFERENT METHODS .IN THIS STUDY, THE STRESS CONCENTRATIONFACTOR (SCF) IN A PLATE WITH A CIRCULAR HOLE UNDER AXIAL TENSION STRESSES WAS INVESTIGATED. THE EMPIRICAL (PETERSON'S) STRESS CONCENTRATION FACTOR (K_T) WAS COMPARED WITH THERESULTS OFANALYTICAL MODEL, REGRESSION ANALYSIS(REGA), FINITE ELEMENT ANALYSIS (FEA), AND ARTIFICIAL NEURAL NETWORK (ANN) MODEL. THE STRESS CONCENTRATION FACTOR (K_T) WAS MODELED USING 5 DIFFERENT METHODS AND THE ACCURACY OF PETERSON'S MODEL WAS TESTED.[3]

S.G. Sarganachari et al.(2019), assess the stress reduction techniques in a uniaxially loaded plate with a Hole [5]. Various stress reduction techniques are available to determine stress concentration factor under different loading conditions such as experimental, analytical and numerical method. Though experimental methods give the most reliable results, it is very costly, as it requires special equipments, testing facilities etc. Analytical solution of every problem is almost impossible because of complex boundary conditions and shapes [4].

E Devaraj et al. (2019) have does numerical and analytical analysis of SCF for a rectangular plate with holes as discontinuities for static loading. They have analysed von-Mises stress for a rectangular plate of dimensions 100 mm x 50 mm x 1 mm with a hole of different radius by analytical and numerical methods.Plate with hole of major radius of 10mm, 8mm and 6mm has been modelled and descritized with SOLID95 elements in ANSYS [5].

Sohanur Rahman (2018) did stress analysis of finite steel plate with a rectangular hole subjected to uniaxial stress using finite element method. In this paper, the effect of von-Mises stress on various types of reinforcements have been investigated and from this analysis it has been observed that single doubler plate gives the most suitable reinforced model. Also author performed the convergence test for the most suitable reinforced model and observe the effect of variation in reinforcement [6].

Olesya Maksymovych et al.(2017) have carried out stress calculation and optimization in composite plates with holes based on the modified integral equation method. They developed numerical method for solving integral equations by the mechanical quadrature method for systems of holes, which takes into account their Eigen-solutions. Precision of the approach and stability of obtained algebraic equations is illustrated in determination of holes shapes with low stress concentration; study of high stress concentration at slit of arbitrary width (additionally used asymptotic method). For a large number of holes with controlled accuracy they calculate the stress [7].

Soni Kumariet al.(2017) carried out stress analysis for an infinite plate with circular holes.Failures such as fatigue cracking and plastic deformation frequently occur at points of stress concentration. In the presented paper, study stress concentration in an infinite isotropic plate around circular hole subjected to transverse, longitudinal and biaxial loading is calculated using analytical approach. For calculating stress concentration around two holes they used complex variable and bipolar coordinate method [8].

Dr. Abdul siddique shaiket al. (2016) carried out finite element analysis (FEA) to do stress concentration analysis of rectangular plate with a hole made with composite material. Finite Element simulations using ANSYS done for stress analysis around the circular hole, made up of different materials. The materials considered are composite material i.e. carbon / epoxy and also with mild steel [9].

A Santos et. al.(2016)carried out simulation of stress concentration factors in combined discontinuities on flat plates. They determined the stress concentration factors for static load conditions (K_c) by using finite elements software. They used ANSYS software for the determination of different parameters flat plates with a central hole subjected to axial load conditions. They also obtained the graphs of K_c for flat plates with combined discontinuities, central hole, groove and central hole, fillet, under axial load conditions [10].

IV.Finite Element Analysis

Numerical analysis for a plate with circular hole geometries subjected to axial tension were carried out on ANSYS 19.0.
The plate with hole is meshed with rectangle elements. The fine mesh is selected to get more accurate results, which in turn resulting in 1650 nodes and 210 elements. Solid Works model and ANSYS meshed model of plate with center hole.

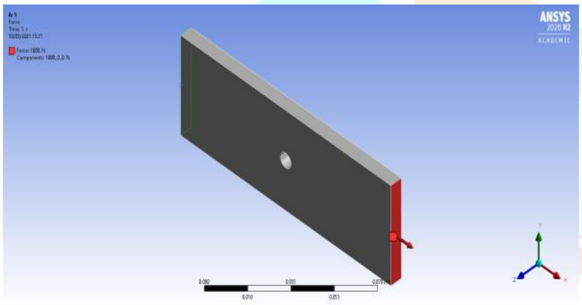


Fig. 3 Loading and boundary conditions

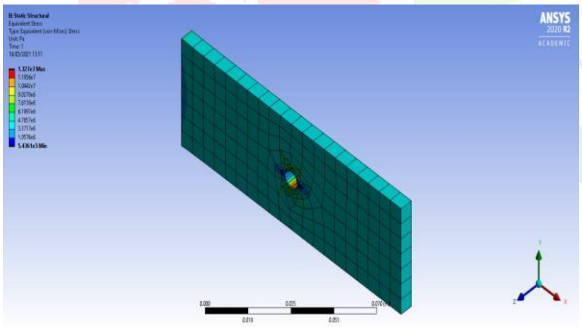


Fig. 4 central hole with axial load

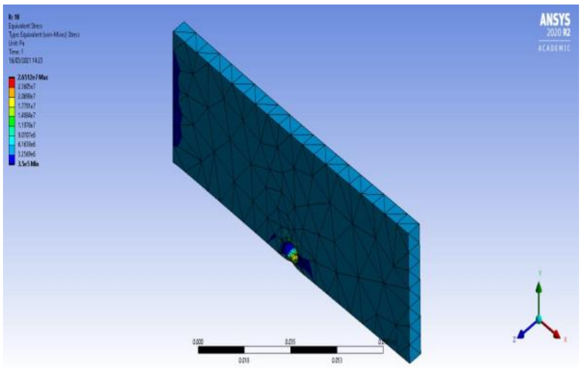


Fig. 5 Near the egde with axial load

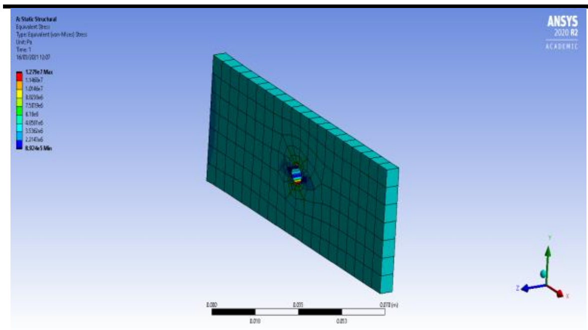


Fig. 6 central hole with compression load

The software used to carry out the finite elements simulation is ANSYS [14]. The developed finite element model is composed of the meshed geometry, the boundary conditions and the loads. Since SCF is independent of mechanical characteristics of the plate, the material chosen for the simulation is the ordinary steel with the conventional elastic mechanical characteristics = 0.3). The mesh is refined to ν(E=210 GPa, carry out at the same time the convergence and the optimization of the data-processing resources in memory and simulation times. All previous studies indicate that the SCF is independent of the applied stress on the structure and only the geometric characteristics are involved in the calculations.

5. Results and Discussion

The stress concentration factor (SCF) is calculated from the maximum stress, which can be determined from the finite element simulation results and the nominal stress.

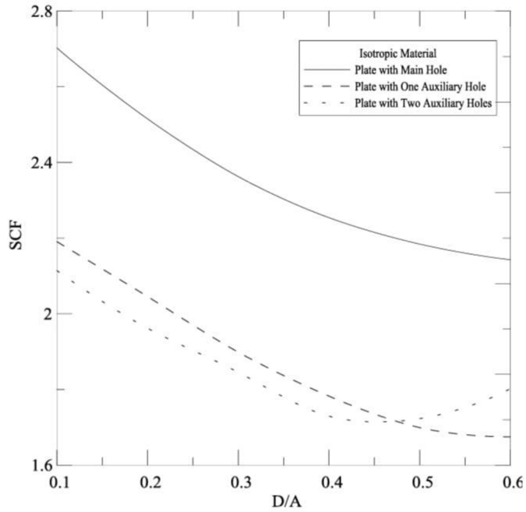
Fig.4 presents the distribution of the normal stress in the applied tensile load with central hole.It is clear that the normal stress is not uniform in the surrounding of the hole. The maximum value of this stress is localised on the bore perpendicular to the applied load direction.

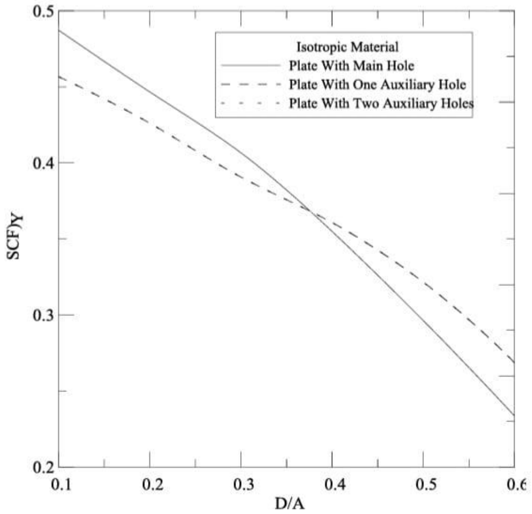
Fig.5 presents the distribution of the normal stress in the applied tensile load with near the edge of the hole.The SCF is maximum value as compared to the central hole.

Fig.6 presents the distribution of the normal stress in the applied compressive load with central hole.The SCF is less than the the central hole with tensile load .

L=Length=120 mm
B= Width=60 mm
D= Hole Diameter=6 mm
e= Dist. from top of the plate to Center of the hole
c= Dist. from bottom of the plate to center of the hole
a= Hole Radius=3 mm
P=Load=1000N
A=Area=204 mm^2

a/c	c/e	e/c	Nominal stress (MPa)	Equivalent (von Mises) stress (MPa	SCF (K_t) based FEA results
0.15	1	1	5	13.45	2.695
1	0.08	12.3	5	26.51	5.408
0.15	1	1	5	12.79	2.609





In general, the maximum stress concentration is always occurred on hole boundary in a finite width plate with central hole under in-plane static loading. The stress concentration factor is maximum at the tip of the hole (perpendicular to loading).

The SCF follows a symmetric trend with respect to D/A ratio in all cases. On the basis of results obtained, it has been seen that the SCF is sensitive to D/A and material properties. The results obtained show that for higher values of D/A, SCF is also higher. SCF reducing as D/A increases. Shear stress increases with introduction of auxiliary holes. Shear stress is maximum with set of auxiliary hole then reduces with two set of auxiliary hole. Deflection in X direction increases with D/A ratio for all the materials.

REFERENCES

1. Shubhrata Nagpa , Dr. S. Sanyal , Dr. N.K.Jain "Analysis And Mitigation Of Stress Concentration Factor Of A Rectangular Isotropic And Orthotropic Plate With Central Circular Hole Subjected To In-Plane Static Loading By Design Optimization" International Journal of Innovative Research in Science, Engineering and Technology Vol. 2, Issue 7, July 2013
2. Pilkey, W.D (1997), Peterson's stress concentration factors, Third edition, (John Wilkey, Newyork).
3. Ozkan, Murat Tolga, Ihsan Toktas, "Determination of the stress concentration factor (Kt) in a rectangular plate with a hole under tensile stress using different methods", Fracture mechanics and modelling, Vol. 58, Issue. 10 , pp. 839-847, (2016).
4. S.G. Sarganachari, Basavaraj R. Endigeri, "Assessment Of Stress Reduction Techniques In An Uni-axially Loaded Plate With A Hole", International Journal of Mechanical and Production Engineering, Vol. 7, Issue. 3 (2019).
5. Devaraj E., P. Sachin Siyal, "Numerical and Analytical Analysis of SCF for A Rectangular Plate with Discontinuities Holes for Static Loading", International Journal of Scientific Research in Multidisciplinary Studies, Vol.5, Issue.12, pp.06-11, December, (2019).
6. Sohanur Rahman, "Stress Analysis of Finite Steel Plate with a Rectangular Hole Subjected to Uniaxial Stress Using Finite Element Method", Marine Science: Research & Development (2018).
7. Olesya Maksymovych, Oleksandr Illiushyn, "Stress calculation and optimization in composite plates with holes based on the modified integral equation method", Engineering Analysis with Boundary Elements, (2017).
8. Soni Kumari, A.K Uphadhyay, K.K Shukla, "Stress Analysis for an infinite plate with circular holes", Materials Today: Proceedings 4 pp. 2323–2332 (2017).
9. Dr. Abdul siddique shaik, Ishrat Meera Mirzana. "Stress Concentration of Rectangular Plate with a Hole Made With Composite Material Using Finite Element Analysis", IOSR Journal of Mechanical and Civil Engineering (IOSR-JMCE) e-ISSN: 2278-1684,p-ISSN: 2320-334X, Volume 13, Issue 4 Ver. VII (2016).
10. A Santos , R Guzman , Z Ramirez and C Cardenas, "Simulation of stress concentration factors in combined discontinuities on flat plates", Journal of Physics: Conference Series 743 (2016).

ISSN: 2278-0181

A Study on various Human Facial Feature Extraction Techniques in High Dimensional Spaces

Jaimin H. Jani, Dr. Subhaschandra Desai

Abstract-In today's era where one's face is used for ease of access for permitted levels of access in either physical or logical-way, it's a very challenging task for the devices equipped with various hardware and software tools to perform such kind of job with desirable accuracy in real time. Feature extraction is a very crucial and important task in facial recognition. In this paper various feature extraction techniques in high dimensional spaces are discussed. The objective of this study is to investigate pattern recognition methods for high-dimensional sample spaces.In a real time scenario and from a performance perspective, the dimensionality could be one of the culprits and makes a significant impact on the effectiveness of the outcome. If the data is transformed to a lower dimensional space by finding a new axis-system in which most of the data variance is preserved in a few dimensions. This reduction may also have a positive effect on the quality of similarity for certain data domains such as text. Our analysis also indicates currently accepted techniques and impact on overall performance as far as the feature extraction phase of facial recognition is concerned.

INTRODUCTION

Face recognition is an active research area with a wide range of applications in the real world. In recent years, a defined face recognition pipeline, consisting of four steps i.e. detection, alignment, representation, and classification has been presented.

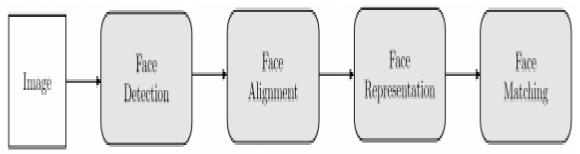


Fig. 1: Face recognition building blocks.

In the detection step the place of the image including face is found. The alignment step ensures the detected face is lined up with a target face or a model. In the representation step the detected face is described in a way that several descriptions with certain aspects about the detected face are presented. Finally, the classification step determines whether a certain feature corresponds with a target face or a model. Face recognition techniques are divided into Geometric and Photometric approaches. Geometric approaches consider individual features such as eyes, nose, mouth and a shape of the head and then develop a face model based on the size and the position of these characteristics. In photometric approaches the statistical values are extracted, subsequently, these values are compared with the related templates. A large number of researches have been devoted to feature extraction based on Gabor filter. A face representation using the Gabor filter, has been of focal importance in the machine vision, image processing and pattern recognition. In face recognition, the feature representation of a face is a critical aspect. If the representation step does not perform well, even the best classifiers cannot produce appropriate results. Good representations are those that on one hand minimize intra-person dissimilarities, on the other hand maximize differences between persons. Additionally, a significant representation should be fast and compact. There are several views related to the classification of the feature extraction methods. One possible classification divides the feature extraction methods into Holistic Methods and Local Feature-based Methods. In the first method the whole face image is applied as an input of the recognition operation similar to the well-known PCA-based method which was used in Kiby and Sirovichfollowed by Turk and Pentland. In the second method local features are extracted, for example the location and local statistics of the eyes, nose and mouth are used in the recognition task. EBGM methods are included in this category. Lades suggested a face recognition system based on DLA (Dynamic Link Architecture) platform, using extracting Gabor jets from each node over the rectangular grid to recognize faces. Wiskottexpanded DLA and introduced EBGM (Elastic Base Graph) method based on a wavelet to recognize the face. However, both LDA and EBGM have a high computational cost. Although the Gabor filters are computationally expensive due to a high dimension of the feature vector the results obtained from them are robust. T.Ojalaintroduced an original LBP operator which is regarded as a strong tool for describing the image texture.

Due to digitization, a huge volume of data is being generated across several sectors such as healthcare, production, sales, IoT devices, Web, organizations. Machine learning algorithms are used to uncover patterns among the attributes of this data. It has been demonstrated that high-dimensional space is significantly different from the three-dimensional (3-D) space, and that our experience in 3-D space tends to mislead our intuition of geometrical and statistical properties in high-dimensional sample spaces.

1.Characteristic Properties of High- Dimensional Spaces

For a fixed number of training samples, increasing the dimensionality of the sample space spreads the data over a greater volume. This process reduces overlap between the classes and enhances the potential for discrimination. Therefore, it is reasonable to expect that high dimensional sample spaces contain more information of capability to detect more classes with more accuracy. However, from the curse of dimensionality, we know that there is a penalty in classification accuracy as the number of features increases beyond some point. Therefore, techniques of carrying out computations at full dimensionality may not deliver the advantages of high-dimensional sample spaces if there are insufficient training samples.

Experiments have shown that high-dimensional sample spaces are mostly empty since data typically concentrate in an outside shell of the sample space far from the origin as the dimensionality increases. This implies that the data samples are usually in a lower dimensional structure. As a consequence, high-dimensional data can be projected to a lower dimensional subspace without losing significant information in terms of separability among the classes by employing some feature extraction techniques. It has been also proved that as the dimensionality of the sample space goes to infinity, lower-dimensional linear projections approach a normality model with a probability approaching one. Here normality implies either a normal or a mixture of normal distributions. It turns out that the normally distributed high-dimensional data concentrate in the tails and uniformly distributed high-dimensional data concentrate in the corners. This makes density estimation task for high-dimensional sample spaces a difficult task. In this case, local neighborhoods become empty, which in turn produces the effect of losing detailed density estimation.

Another interesting observation was related to the first and the second order statistics of data samples. It has been shown that for low-dimensional sample spaces, class means representing first order statistics play a more important role in discriminating between classes than the class covariances representing second order statistics. However, as dimensionality increases, class covariance differences become more important.

In summary, the dimensionality of the sample space must be reduced before the application of the classifier to data samples in high-dimensional sample spaces. However, in order to keep the discriminatory information, which the high-dimensional sample spaces provide, good dimension reduction techniques are needed. In this study, the dimension reduction techniques for high-dimensional sample spaces are investigated.

2.DIMENSIONALITY REDUCTION

Dimensionality reduction usually improves the accuracy of recognition of a pattern recognition system besides saving memory and time consumptions. This seems somewhat paradoxical since dimensionality reduction usually reduces the information content of the input data. However, a good dimensionality reduction technique keeps the features with the high discriminative information and discards the features with redundant information. Thus, the worst effects of the curse of dimensionality are reduced after the dimensionality reduction process, and often improved performance is achieved over the application of the selected classifier in the original sample space. But given a set of features, how can the best set of features for classification be selected? Given a set of features, selection of the best set of features can be achieved in two different ways.

The first approach is to identify the features that contribute most to class separability. Therefore, our task is the selection of previously decided features out of our initial d features. This is called feature selection. The second approach is to compute a transformation which will map the original input space to a lower-dimensional space by keeping the most of the discriminative information. This transformation can be linear or nonlinear combinations of the samples in the training set. This approach is usually called the feature extraction. Both approaches require a criterion function, J, which is used to judge whether one subset of features is better than another. Exploring high-dimensional data is central to many application domains such as statistics, data science, machine learning, and information visualization. The main difficulty encountered in this task is the large size of such datasets, both in the number of observations (also called samples) and measurements recorded per observation (also called dimensions, features, variables, or attributes)

FEATURE SELECTION

In this approach we select the best set of features for classification out of original d features. We must first define a criterion function, J, to accomplish this task. The selected criterion is evaluated for all possible combinations of features systematically selected from d features. Then, we select the set of features for which the criterion is maximum as our final features. However, this task is not very straightforward because there are

$$\frac{d!}{(d-\tilde{d})!\,\tilde{d}!}$$

possible combinations for evaluation. As a consequence, this procedure may not be feasible even for moderate values of d and therefore, we will not consider the feature selection methods in this study since we are only interested in the data sets with high-dimensional spaces.

3. FEATURE EXTRACTION

In this approach we seek a transformation which will map the original input space to a lower dimensional space by keeping the features offering high classification power. The optimization is evaluated over all possible transformations of the data samples. Let denote the sought transformation for which, where ϖ is the family of allowable transformations and x refers to the training set samples. The new samples in the transformed space are computed by $y = W(x)$. The criterion function is typically a measure of distance or similarity between training set samples.

Linear Feature Extraction Methods

Feature extraction has been one of the most important issues of pattern recognition. Most of the feature extraction literature has centered on finding linear transformations, which map the original high-dimensional sample space into a lower-dimensional space that hopefully contains all discriminatory information. As explained previously, the principal motivation behind dimensionality reduction by feature extraction is that it may reduce the worst effects of the curse of dimensionality. Also linear feature extractions techniques are often used as pre-processors before more complex nonlinear classifiers. In the following sections we discuss these linear methods.

Generally, the face recognition process is divided into 3 regions such as Holistic method use the original image as an input for the face recognition system. The examples for holistic methods are PCA, LDA, and ICA and so on. In the Feature based method, the local feature points such as eye, nose, and mouth are first extracted, then it will be sent to the classifier. Finally, a Hybrid method is used to recognize both the local feature and whole face region. In Dimensionality reduction, Feature extraction is an important task to collect the set of features from an image. According to the author, Feature extraction or transformation is a process through which a new set of features is created. The feature transformation may be a linear or nonlinear combination of original features. This survey provides some of the important linear and nonlinear techniques listed as follows.

3.1 Principal Component Analysis (PCA)

PCA is one of the popular techniques for both dimensionality reduction and face recognition since the 1990's. Eigenfacesbuilt on the PCA technique is introduced by M.A.Turk and A.P.Pentland. It is a holistic approach where the input image is directly used for the process. PCA algorithm can be used to find a subspace whose basis vectors correspond to the maximum variance directions in the original *n* dimensional space. PCA subspace can be used for presentation of data with minimum error in reconstruction of original data. More survey papers provide the information for PCA techniques. MPCA and KPCA are fully based on the PCA technique.

3.2 Linear Discriminant Analysis (LDA)

LDA is one of the most famous linear techniques for dimensionality reduction and data classification. The main goal of the LDA consists in finding a base of vectors providing the best discrimination among the classes, trying to maximize the between-class differences, minimizing the within-class ones by using scatter matrices. It also suffers from a small sample size problem which exists in high dimensional pattern recognition tasks where number of available samples are smaller than dimensionality of the samples. DLDA, R-LDA, and KDDA are variations of LDA. This technique is also discussed in more survey papers.

3.3 Singular Value Decomposition (SVD)

SVD is an important factor in the field of signal processing and statistics. it is the best linear dimensionality reduction technique based on the covariance matrix. The main aim is to reduce the dimension of the data by finding a few orthogonal linear combinations of the original variables with the largest variance. Most of the researchers have also used this technique for face recognition.

3.4 Independent Component Analysis (ICA)

ICA is a statistical and computational technique for enlightening the hidden factors that underlie sets or random variables, measurements, or signals. ICA is superficially related to principal component analysis and factors analysis. The ICA algorithm aims at finding S components as independent as possible so that the set of observed signals can be expressed as a linear combination of statistically independent components. It use cosine measures to perform the covariance matrix and also it is better than the PCA and LDA performance.

3.5 locality Preserving Projections (LPP)

LPP can be seen as an alternative to Principal Component Analysis (PCA). When the high dimensional data lies on a low dimensional manifold embedded in the ambient space, the Locality Preserving Projections are obtained by finding the optimal linear approximations to theEigen functions of the Laplace Beltrami operator on the manifold. As a result, LPP shares many of the data representation properties of nonlinear techniques such as Laplacian Eigenmapsor Locally Linear Embedding.

3.6 multi Dimensional Scaling (MDS)

Multidimensional Scaling (MDS) is a linear Model for dimensionality reduction. MDS generates low dimensional codes placing emphasis on preserving the pairwise distances between the data points. If

the rows and the columns of the data matrix D both have mean zero, the projection produced by MDS will be the same as that produced by PCA. Thus, MDS is a linear Model for dimensionality reduction having the same limitations as PCA.

3.7 partial Least Squares

Partial least squares is a classical statistical learning method. It is widely used in chemo metrics and bioinformatics etc. In recent years, it is also applied in face recognition and human detection. It can avoid the small sample size problem in linear discriminant analysis (LDA). Therefore it is used as an alternative method of LDA.

4. NON LINEAR FEATURE EXTRACTION OF DIMENSIONALITY REDUCTION TECHNIQUES

This section presents a general introduction to nonlinear feature extraction methods employing kernel functions. The kernel trick concept has been introduced here, and this trick is applied to the linear DCV Method to make it a nonlinear method.

Non-linear methods can be broadly classified into two groups: a mapping (either from the high dimensional space to the low dimensional embedding or vice versa), it can be viewed as a preliminary feature extraction step and visualization is based on neighbor’s data such as distance measurements. Research on non-linear dimensionality reduction methods has been explored extensively in the last few years.

4.1 An Introduction to Kernel Feature Extraction Methods

Sometimes linear methods may not provide sufficient nonlinear discriminant power for classification of linearly non-separable classes (e.g., exclusive-or problem). Thus, kernel methods have been proposed to overcome this limitation. The basic idea of these methods is first to transform the data samples into a higher-dimensional space $\Im$ via nonlinear mapping φ (.), and then apply the linear methods in this space. More formally, we apply the mapping

$\varphi : Rd \rightarrow \Im,\ x\ a \varphi\ (x)$

to all the data samples. The motivation behind this process is to transform linearly non-separable data samples into a higher-dimensional space where the data samples are linearly separable as illustrated in Figure 4.1. Since the mapped space is nonlinearly related to the original sample space, nonlinear decision boundaries between classes can be obtained for classification. This approach seems to contradict the curse of dimensionality phenomenon since it increases the dimensionality of the sample space for a fixed number of available training set samples. A satisfactory explanation for this dilemma lies in statistical learning theory. This theory tells us that learning in high-dimensional space can be simpler if one uses low complexity, i.e., a simple class of decision rules such as linear classifiers. In other words, it is not the dimensionality but the complexity of the function that matters. In some recognition tasks we may have sufficient knowledge about the problem and can choose φ (.) by hand. If the mapping is not too complex and is not too high-dimensional, we can explicitly apply this mapping as happens in Radial Basis Networks or Boosting Algorithms. However, in most cases we may not have sufficient prior knowledge to design φ (.), or the mapping of the data samples into a higher-dimensional space explicitly cannot be intractable. In such cases, we utilize kernel functions to circumvent these limitations.

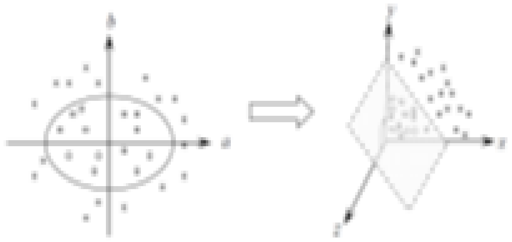


Figure 2: Kernel (nonlinear) mapping of 2-dimensional data into 3-dimensional space by polynomial kernel function.

In the following, a brief introduction to several non-linear dimensionality reduction techniques will be given.

4.1.1 Kernel Principal Component Analysis (KPCA)

Kernel PCA (KPCA) is the reformulation of traditional linear PCA in a high-dimensional space that is constructed using a kernel function. In recent years, the reformulation of linear techniques using the 'kernel trick' has led to the proposal of successful techniques such as kernel ridge regression and Support Vector Machines. Kernel PCA computes the principal eigenvectors of the kernel matrix, rather than those of the covariance matrix. The reformulation of traditional PCA in kernel space is straightforward, since a kernel matrix is similar to the in product of the data points in the high-dimensional space that is constructed using the kernel function. The application of PCA in kernel space provides Kernel PCA the property of constructing nonlinear mappings.

4.1.2 Isometric Mapping (ISOMAP)

Most of the linear methods do not take the neighboring data points into an account. ISOMAP is a technique that resolves this problem by attempting to preserve pair wise geodesic (or curvilinear) distances between data points. The approximation of geodesic distance is divided into two cases. For neighboring points, Euclidean distance in the input space provides a good approximation to geodesic distance and faraway points, geodesic distance can be approximated by adding up a sequence of "short hops" between neighboring points. ISOMAP shares some advantages with PCA, LDA, and MDS, such as computational efficiency and asymptotic convergence guarantees, but with more flexibility to learn a broad class of nonlinear manifolds.

4.1.3 Locally Linear Embedding

Locally linear embedding (LLE) is another approach which addresses the problem of nonlinear dimensionality reduction by computing low dimensional, neighborhood preserving embedding of high-dimensional data. It is a technique that is similar to ISOMAP in that it also constructs a graph representation of the data points. It describes the local properties of the manifold around a data point x_i by writing the data point as a linear combination w_i (the so-called reconstruction weights) of its k nearest neighbors x_{ij} and attempts to retain the reconstruction weights in the linear combinations as good as possible.

4.1.4 Laplacian Eigenmaps:

A closely related approach to locally linear embedding is Laplacian eigenmaps. Given *t* points in *n*-dimensional space, the Laplacian eigenmaps Method (LEM) starts by constructing a weighted graph with *t* nodes and a set of edges connecting neighboring points. Similar to LLE, the neighborhood graph can be constructed by finding the *k* nearest neighbors. The final objectives for both LEM and LLE have the same form and differ only in how the matrix *is* constructed.

4.1.5 Stochastic Neighbor Embedding:

Stochastic Neighbor Embedding (SNE) is a probabilistic approach that maps high dimensional data points into a low dimensional subspace in a way that preserves the relative distances to near neighbors. In SNE, similar objects in the high dimensional space will be put nearby in the low dimensional space, and dissimilar objects in the high dimensional space will usually be put far apart in the low dimensional space. A Gaussian distribution centered on a point in the high dimensional space is used to define the probability distribution that the data point chooses other data points as its neighbors. SNE is superior to LLE in keeping the relative distances between every two data points.

4.1.6 Semi Definite Embedding (SDE):

Semi definite Embedding (SDE), can be seen as a variation of KPCA and an algorithm is based on semi definite programming. SDE learns a kernel matrix by maximizing the variance in feature space while preserving the distances and angles between nearest neighbors. It has several interesting properties: the main optimization is convex and guaranteed to preserve certain aspects of the local geometry; the method always yields a semi positive definite kernel matrix; the eigenspectrum of the kernel matrix provides an estimate of the underlying manifold's dimensionality; also, the method does not rely on estimating geodesic distances between far away points on the manifold. This particular combination of advantages appears unique to SDE.

5. CONCLUSION

Because of varying applications and span over different domains, selection of appropriate feature extraction techniques make a major impact in computation required (i.e. time and space complexity) in face recognition. Scholars have conducted and explored various aspects vigorously in this area for the past many years, and though significant amounts of progress has been achieved so far. Feature extraction is one of the most preprocessing and fundamental task in face recognition tasks. This paper contained a detailed survey on various existing feature extraction techniques for face recognition. Different face recognition algorithms can be applied on available databases. Even when the same database is used, researchers may use different protocols for testing. After a detailed review of a number of research papers, we found two main points (1) For the best-performing supervised defect prediction models, correlation and consistency-based feature selection techniques should be appropriate and (2) Neural network-based feature reduction techniques generate features that have a small variance across both supervised and unsupervised defect prediction models. In summary, a face recognition system should not only be able to cope with variations in illumination, expression and pose, but also recognize a face in real-time. We recommend that practitioners who do not wish to choose a best-performing defect prediction model for their data use a neural network-based feature reduction technique.

REFERENCES

[1] Ngoc-Son Vu, H. M. Dee and A. Caplier, (2012) "Face recognition using the POEM descriptor", Pattern Recognition.
[2] C. Liu and H. Welchsler, (2001) "Gabor feature classifier for face recognition", in processing of the ICCV, Vol. 2, No. 5, pp 270-275.
[3] J.R. Movellan, "Tutorial on Gabor filters",http://mplab.ucsd.edu/tutorials/gabor.pdf.
[4] M. Zhou, and H. Wei, (2006) "Face verification using Gabor Wavelets and AdaBoost", 18th International Conference on Pattern Recognition, pp 404-407.
[5] M.Kirby and L. Sirovish, (1990) "Application of the Karhunen-Lo_ve procedure for the characterization of human faces", IEEE Transactions on Pattern Analysis and Machine Intelligence12, pp 103-108.
[6] M.Turk and A.P. Pentland, (1991) "Eigen faces for recognition", Journal of Cognitive Neuroscience, pp 71-86.
[7] C. Aguerrebere, G. Capdehourat, M. Delbracio, M. Mateu, A. Fern´andez and F. Lecumberry, (2007) "Aguar´a: An Improved Face Recognition Algorithm through Gabor Filter Adaptation", Automatic Identification Advanced Technologies.
[8] M.Lades, J.C.Vorbruggen, J.Buhmann, J.Lang, C.V.Malsburg, C.Wurtz and W.Konen, (1993) "Distortion invariant object recognition in tha dynamic link architecture", IEEE Trans.Computers, Vol.42, No.3, pp 300-311.
[9] L.Wiskott, J.M.Fellous, N.Kruger, and C.VMalsburg, (1997) "Face recognition by elastic bunch graph matching, IEEE Trans, Pattern Aal. Match.Intel., Vol.19, No.7, pp 775-779.
[10] A. Bayesian, and C.H. Liu,(2007) "On Face Recognition using Gabor Filters", World Academy of Science Engineering and Technology 28, pp 51-56.
[11] T. Ojala, Pietikinen and Menp, (2002) "Multi resolution gray-scale and rotation invariant texture classification with local binary patterns", IEEE Transaction on Pattern Analysis and Machine Intelligence, pp 971-987.
[12] Jimenez, L. O. and Landgrebe, D. A. (1998) Supervised classification in high dimensional space: geometrical, statistical, and asymptotical properties of multivariate data. IEEE Transactions on Systems, Man, and Cybernetics-Part C: Applications and Reviews, 28(1), 39-54.
[13] Veerabhadrappa, LalithaRangarajan," Bi-level dimensionality reduction methods using feature selection and feature extraction" International Journal of Computer Applications (0975 – 8887) Volume 4 – No.2, July 2010.
[14] Rama Chellappa, Charles L. Wilson, And SaadSirohey,"Human and machine recognition of faces: A Survey",Proceedings of the IEEE,1995
[15] William A. Barrett," A Survey of Face Recognition Algorithms and Testing Results",Proceedings of the IEEE,1998.
[16] W.Zhao,R.Chellapa,A.Rosenfield,P.J.Philips," Face Recognition : A Literature Survey",2001
[17] W.Zhao,R.Chellapa,A.Rosenfield,P. J.Philips," Face Recognition : A Literature Survey",ACM proceedings,2003
[18] XiaoyangTana,b, SongcanChena,c,∗, Zhi-Hua Zhoub, FuyanZhangb," Face recognition from a single image per person:Asurvey",Published in Elseiver,2006
[19] Patil A.M., Kolhe S.R. and Patil P.M," 2D Face Recognition Techniques: A Survey",2010
[20] M. Turk and A. Pentland, "Eigenfaces for recognition", Journal of Cognitive Neuroscience, vol. 3, No. 1, 1991, pp.71 - 86.
[21] S.K.Sandhu, SumitBudhiraja," Combination of Nonlinear Dimensionality Reduction Techniques for Face Recognition System",published in IJERA
[22] S.Sakthivel," enhancing face recognition using improved dimensionality reduction and feature extraction algorithms –an evaluation with orl database" international journal of engineering science and technology,2010
[23] Shylaja S S, K N Balasubramanya Murthy and S Natarajan," Dimensionality Reduction Techniques for Face Recognition", (IJACSA) International Journal of Advanced Computer Science and Applications, 2011
[24] Yunfei Jiang and Ping Guo," Comparative Studies of Feature Extraction Methods with Application to Face Recognition"IEEE,2007
[25] Ion Marqu´es," Face Recognition Algorithms",2010.
[26] CHEN Cai-ming, Zhang Shi-qing,ChenYuefen," Face Recognition Based on MPCA", 2nd International Conference on Industrial Mechatronics and Automation,2010
[27] Weilin Huang and Hujun Yin," linear and nonlinear dimensionality reduction for face recognition",IEEE,2009
[28] Schölkopf, B. and Smola, A. J. (2002) Learning with Kernels. MIT Press.
[29] Müller, K.-R., Mika, S., Rätsch, G., Tsuda, K. and Schölkopf, B. (2001) An introduction to kernel-based learning algorithms. IEEE Transaction on Neural Networks, 12, 181-201.
[30] Ali Ghodsi," Dimensionality Reduction A Short Tutorial",2006
[31] Renqiang Min," A Non-linear Dimensionality Reduction Method for Improving Nearest Neighbour Classification",2005
[32] Thippa Reddy Gadekallu,Praveen Kumar Reddy,KuruvaLakshman,RajeshKaluri, "Analysis of Dimensionality Reduction Techniques on Big Data",2020
[33] MateusEspadoto, Rafael M. Martins, Andreas Kerren, Nina S. T. Hirata, and Alexandru C. Telea, "Towards a Quantitative Survey of Dimension Reduction Techniques",2019

A Study to Develop Human Face Recognition using PCA, Neural Networks and Wavelet

Jaimin H Jani, Dr. Subhash Chandra Desai

***Abstract*-A method to improve the precision of the face recognition with the help of an integrating of WT, PCA and neural networks has been presented in this paper. The three main critical issues for face recognition are- preprocessing, feature extraction and classifying rules. A hybrid approach for employing the three issues has been presented in this paper. A combination of wavelet transform (WT) and PCA has been used for preprocessing and feature extraction. Neural Network is discussed for achieving a fast decision in the presenceofvarietyoffacialexpressions,intheclassificationstage.Overall,improvisationsintheproposed method's accuracy isdone.**

INTRODUCTION

Theauthenticationofusershavebeenincreasinginthepastfewyears,becausetherequirementofsecurity isomnipresent.originally,identificationcardsandpasswordswerepopularforprovingauthenticity,though security through these methods is not very reliable. The latest interest of the researchers are authentication technologies based on biology, such as the ones that use iris, fingerprint, face, print of the palm and voice. Face recognition has gained popularity, largely because the process of authentication is done in a hands free way, without interrupting the activity of the user in any way. also, it is economic due to the low cost of the cameras and computer. Psycho-physicists and neuroscientist have focused on issues like face uniqueness, organization of face memory and the perception of faces by infants in the past 20 years. At the same time, engineers have studied, developed and designed algorithms of face recognition in the last 20 years. This paper focusses on the work of the engineers. Content based approach and face based approach are the two approached of face recognition system done by the computers.

There lation ship between the face boundary and facial eatures like nose, eyes, mouth are used in the content based approach. A huge classification error can be committed in the process of derivation, since all the human faces have features that are similar.

Inthefacebasedapproach,thefaceiscapturedasawholeandistreatedasa2Dpattern.Thefaceismatched with the statistical regularities. Principal Component Analysis (PCA) is a face based approach, which has been proven to beeffective.

Karhunen- Loeve (KL) transform was proposed the representation of human faces by Sirovich and Kirby. The faces are represented with the help of eigenfaces, which are the linear combinations of weighed eigenvector, in this method. However, a system of face recognition that makes use of the PCA has been developed by Turk and Pentland. But, this method is not free from limitations. The two limitations of this methodsarelargeloadofcomputationandpoor powerofdiscrimination.Themeasuredsimilaritybetween 2 pictures of the same individual by using the PCA method is high. However, the measured similarity of 2 pictures of different people is also high. Therefore, the discrimination power of this PCA method is very poor.

This drawback of PCA was improved by addition of Linear Discriminant Analysis (LDA) by Swets and Weng. A different method for selection of eigenfaces was suggested by O'Toole et al. they stated that the eigenvectors which have large eigenvalues is not the best method to differentiate face image. It was also presented by them that the representations of low dimension are efficient in identification of physical features of the face, like race and gender even though they might not be the best way for the recognition of human faces.

HeavyloadofcomputationintheprocessoffindingeigenvectorsisanotherprobleminPCAbasedmethod is the h. The typical value of computational complexity of $O(d^2)$ is 128x128, where d= number of pixels. The cost of computation is beyond many existing computer's power. However, Matrix Theory tells us that if the number of training images (N) is smaller than the value of d, the complexity of computation will be decreased to $O(N^2)$. Then also if N will increase the load of the computation is increased in cubic order. A newapproachintheapplicationofPCAinthelightofthealreadyexistingPCAapproachhasbeenproposed here. It is proposed that in this method, the image is decayed into many sub bands using the wavelet transform with many frequencycomponents.

Theresultshaveshownthatthe3 levelwavelethasperformedwellinfacerecognition.Themethodwhich hasbeenproposedinthispaperdoesn'tworkontheimageresolutionof128x128,butonalowerresolution of 16x16. Hence, the computational complexity is reduced significantly for many applications, where the training images are more than 16x16. increASed accuracy in the recognition and better discrimination power was observed when PCA was applied on wavelet transform (WT) than when PCA was applied on the entire of the original image.

REVIEW OF PCA

Some major details of PCA are as follows:

Let $X = \{X_{n,} n = 1,..., N\} \in R$ be an ensemble of vectors. When dxd is the product of width and height of the image, the row concatenation of the data of the image is form in the applications of imaging.

1^{N}

Let be the average vector in the ensemble $E(X) = X$ N n=1fter subtracting the average from each elementof X,

ensemble of vectors, $X = \{X_n, n = 1,..., N\}$ with $X_n = X_n - E(X)$ is received ·

covariance matrix M for the ensemble X is defined by $M = \mathrm{cov}(X) = E(X \otimes X)$, Where M is $d^2 xd^2$ matrix, with elements.

It is a well-known fact of the matrix theory that matrix M is always positive and will only have eigenvaluesthatarenon-negative.ThematrixM oftheeigenvectorsformabasisforRdxd.Thisbasis is called K-L basis.

The eigenvectors in K are arranged in a descending order of eigenvalues in many applications. In order to compute the dxd eigenvalue from M, 2xd2 matrix has to be solved. In most chances, d=128, hence 16x16 matrix is solved for calculating the eigenvectors and eigenvalues.

Computer system's requirement for the memory and computation are very high. Matrix theory states that ifN<dxd,thatisN,whichisthenumberoftrainingimage,issmallerthanM,thecomputationalcomplexity decreases to O(N). Therefore, the implementation of PCA in characterizing the faces has become flexible. The number of training images in most researches is around 200. But the M rises when the total number of training images in huge, such as2000.

An image's Wavelet decomposition

In last 10 years, WT has become a useful tool in the analysis of the image. In this paper, WT has been chosen as the option of choice for image decomposition because-

- The resolution of the sub images are decreased when an image is decomposed by using the WT method. so, the computational complexity also decreases because it operates on an image which has alow resolution. It was observed by Harmon that a resolution of 16x16 is enough for human face recognition. When compared to the original image having a resolution 128x128, it leads to a reduction off the sub image by 64 times, thus implying a reduction on the computational load of 64 times.
- TheimagesaredecomposedintosubbandswhichcorrespondtovariousfrequencyrangesundertheWTmethod. The computational overhead is minimized as the sub bands readily meet the input requirement in the system proposed in this paper.
- While the Fourier decomposition support only the global information in the frequency domain, the WTmethodofdecompositionofimagesprovideslocalinformationindomainsfor bothfrequencyandspace.

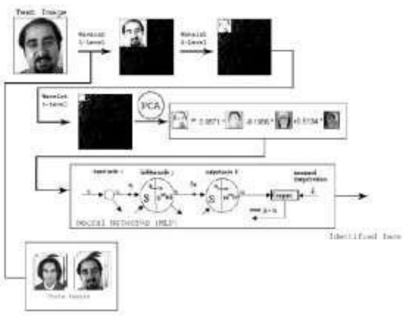
a modified

In this paper, we applied two well-known mother wavelet Daubechies and Haar. We proposed method that uses by coefficients:

h0 = 0.48296291314453 h1 =0.83651630373781
h2 =0.22414386804201
h3 =0.12940952255126
For daubechies mother wavelet and coefficients:
h0 =0.5,h1 =0.5

PROPOSED METHOD

In order to overcome the limitations of the PCA method, this wavelet based PCA method has been developed. also, utilizationof neural networks have been used for classifying faces. A multilayer architecture was adopted,which is fed by the vectors formed by combining wavelet and PCA and decreased inputunits.UsingaparticularfrequencybandofanimageofthefaceforPCAforsolvingthefirstproblem of PCA has beenproposed.



Using a reduced resolution image for dealing with the second limitation of the PCA has been proposed. The proposed system has two stages. one, "training step" in which extractions of features, reduction of dimensions and adjustment of weight of MLP neural networks is done. Second, "recognition step" for identifying the unknown images of faces.

In the "training stage", "feature extraction of reference images" and "adjusting the neural network parameters" is included. In interested domain, the "representational basis" of the images is identified in featureextraction.Then,inputimageistranslatedinaccordancewiththerepresentationalbasis(whichhave been identified in the training stage) in the "recognitionstage".

The 3 important steps in the "training stage" are-

1. For decomposing the reference images, WT is applied. Then, by the decomposition of the wavelet in three levels, sub images of 16x16 pixels which have been obtained, areselected.
2. For obtaining a set of representational basis, by selecting d' eigenvectors which correspond to large eigenvalues and sub space projection, PCA is applied on the subimages.
3. The obtained features of the reference images in the precious step are then used for training neural networks with the help of propagation algorithm. The processing carried out in both the training and recognition stage is similar, the only difference being in the recognition stage, the input unknown images are matched with reference images in the recognition stage. WT and PCA are used for transforming the unknown face images intro the representational basis when an unknown face is presented in the recognitionstage.

EXPERIMENTAL RESULT

The database of face image of Yale university and face database of ORL is used for evaluating the the method that has been proposed in this paper.

All the images in the database of Yale university have a 160x121 resolution. But the WT can't be applied as the images' dimension are not the power of 2. The images were then cropped to 91x91, and hen resized in 128x128. A third level of WT decomposition was use for changing the resolution of images.

Table 1 and table 2 show the results of the proposed algorithm on the database of Yale university and database of ORL, and have used the hair mother wavelet. The results of the proposed algorithm in the database of Yale and ORL which have used Daubechies mother wavelet has been shown in table 3 and 4. The performance in recognition on the test image of the Yale and ORL database which have used various components have been shown in table 5 and 6.

TABLE 1. Algorithm applied on Yale database and haarmotherwavelet.

	PCA on image	PCA on LL band of three level wavelet
Size of image	128*128	16*16
Recognition rate	81.78%	82.2%

TABLE 2. Algorithm applied on ORL database and haarmotherwavelet.

	PCA on image	PCA on LL band of three level wavelet
Size of image	128*128	16*16
Recognition rate	90%	91.80%

TABLE 3. Algorithm applied on ORL database and Daubechies motherwavelet.

	PCA on image	PCA on LL band of three level wavelet
Size of image	128*128	16*16
Recognition rate	90%	97.68%

TABLE 4. Algorithm applied on Yale database and Daubechiesmotherwavelet.

	PCA	PCA on LL band of three level wavelet
Size of image	128*128	16*16
Recognition rate	81.78%	90.35%

TABLE 5. Recognition performance on testimagesof Yale database using the number ofprincipal components.

Number	ANN structure	Recog.	Average
1-15	15:25:15	88.37%	86.56%
1-25	25:30:15	90.35%	89.23%
1-35	35:30:15	89.78%	87.24%
1-45	45:25:15	88.92%	87.68%
1-60	60:35:15	83.78%	88.23%
1-80	80:40:15	85.56%	
1-105	105:45:15	84.76%	83.67%

TABLE 6. Recognition performance on test images of ORL database using the number of principal components.

Number of P.C	ANN structure	Recog. rate (15 attempts)	Average of recognition
1-25	25:40:40	95.37%	94.14%
1-30	30:80:40	94.47%	93.15%
1-35	35:80:40	96.81%	95.45%
1-40	40:40:40	97.68%	96.56%
1-50	50:40:40	96.56%	95.24%
1-100	100:60:40	92.22%	91.72%

TABLE 7. Recognition performance on test images of Yale database using MLP Neural networks by 25of principal components.

Number of P.C	ANN structure	Recog.rate(10 attempts)	Average of recognition
1-25	25:15:15	89.69%	87.56%
1-25	25:20:15	90.06%	89.45%
1-25	25:25:15	90.10%	89.25%
1-25	25:30:15	90.35%	89.23%
1-25	25:40:15	90.05%	87.45%
1-25	25:50:15	90.%	88.34%
1-25	25:60:15	89.86%	87.24%

TABLE 8. Recognition performance on test images of ORL database using MLP Neural networks by 40 of principal components.

Numberof P.C	ANN structure	Recog. Rate (15attempts)	Averageof recognition
1-40	40:10:40	89.67%	87.64%
1-40	40:20:40	91.34%	90.78%
1-40	40:30:40	96.99%	94.67%
1-40	40:40:40	97.68%	96.58%
1-40	40:50:40	96.89%	95.24%
1-40	40:60:40	96.57%	95.67%
1-40	40:70:40	95.98%	94.67%

CONCLUSION

A hybrid approach for face recognition has been presented in this paper, by taking care of three issues. For stagesoffeaturerecognitionandpreprocessing, WTandPCAhavebeenappliedina combinedform.And MLP has been explored for quick decision making when there is a wide variety of facial variations in the classification stage. It can be concluded based on the experiments done on Yale university and ORL databasethatacombinationofWT,PCAandMLPyieldsmost favorableperformance,becauseitexhibits the lowest redundant rate, lowest training time and highest rates ofrecognition.

The proposed method also exhibits the a low load of computation in both the stages- training and recognition.

REFERENCES

[1] R. Chellappa, C. L. Wilson and S. Sirohey, "Human and machine recognition of faces: a survey,"
[2] Proceedings of the IEEE, May 1995, Vol. 83, No. 5, 705-740.
[3] G. Chow and X. Li, "Towards a system for automatic facial feature detection," Pattern
[4] Recognition, 1993, Vol. 26, No. 12, 1739-1755.
[5] F. Goudail, E. Lange, T. Iwamoto, K. Kyuma and N. Otsu, "Face recognition system using local autocorrelations and multiscale integration," IEEE Trans. PAMI, 1996, Vol. 18, No. 10, 1024-1028.
[6] K. M. Lam and H. Yan, "Locating and extracting the eye in human face images", Pattern Recognition, 1996, Vol. 29, No.5 771-779.
[7] D. Valentin, H. Abdi, A. J. O'Toole and G. W. Cottrell, "Connectionist models of face processing: A Survey," Pattern Recognition, 1994, Vol. 27, 1209-1230.
[8] A. L. Yuille, P. W. Hallinan and D. S. Cohen," Feature extraction from faces using deformable templates," Int. J. of Computer Vision, 1992, Vol. 8, No. 2, 99-111.
[9] M. Kirby and L. Sirovich," Application of the Karhunen- Loeve procedure for the characterization of human faces" IEEE Trans. PAMI., 1990, Vol. 12, 103-108.M. Turk and A. Pentland," Eigenfaces for recognition,"J. Cognitive Neuroscience, 1991, Vol. 3, 71-86.
[10] M. V. Wickerhauser, Large-rank "approximate component analysis with wavelets for signal feature discrimination and the inversion of complicated maps," J. Chemical Information and Computer Sciences, 1994, Vol. 34, No. 5, 1036-1046.
[11] L.Sirovich and M. Kirby, "Low-dimensional procedure for the characterization of human faces," J. Opt. Soc. Am. A, 1987, Vol. 4, No. 3, 519-524.
[12] D. L. Swets and J. J. Weng, "Using discriminant eigenfeatures for image retrieval," IEEE Trans.PAMI, 1996, Vol. 18, No. 8, 831-836.
[13] A. J. O'Toole, H. Abdi, K. A. Deffenbacher and D. Valentin, "A low-dimensional representation of faces in the higher dimensions of the space, "J. Opt. Soc. Am., A, 1993, Vol. 10, 405-411.
[14] A. K. Jain," Fundamentals of digital image processing,"Prentice Hall, 1989, pp.163-175.
[15] A.I. Wasserman."Neural Computing: Theory and Practice " New York: Van Nostrand Reinhold,
[16] 1989.
[17] Golovko V., Gladyschuk V." Recirculation Neural Network Training for Image Processing,"
[18] Advanced Computer Systems, 1999, P. 73-78.
[19] L. Harmon, "The recognition of faces," Scientific American, 1973, Vol. 229, 71-82.
[20] I. Daubechies, "Ten Lectures on Wavelets, CBMS-NSF series in Applied Mathematics," SIAM Press, Philadelphia, 1992, Vol. 61.
[21] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis," IEEE Trans.
[22] Information Theory, 1990, Vol. 36, No. 5, 961-1005.
[23] A. Pentland, B. Moghaddam and T. Starner," View- based and modular eigenspaces for face recognition," Proc. IEEE Conf. Computer vision and Pattern Recognition, 1994, Seattle, June, 84-91.
[24] H. A. Rowley, S. Baluja and T. Kanade, "Neural network-based face detection," IEEE Transaction
[25] on PAMI, 1998, Vol. 20, No. 1,23-38.
[26] E.M.-Tzanakou, E. Uyeda, R. Ray, A Sharma, R. Ramanujan and J. Dong, "Comparison of neural network algorithm for face recognition," Simulation, 1995, 64, 1, 15- 27.
[27] D. Valentin, H. Abdi and A. J. O'Toole, "Principal component and neural network analyses of face images: Explorations into the nature of information available for classifying faces by sex,", Progress in mathematical psychology, 1996, Hillsdale: Erlbaum, in press.
[28] K.Fukunaga, "introduction to Statistical Pattern Recognition," Academic press. S.Z. Li and J. Lu, "Face Recognition Using the Nearest Feature Line Method, "IEEE Trans. Neural Networks, . 1999, vol. 10, no.2,pp .439-443, mar. Y. Meyer, "Wavelets: Algorithms and Applications," SIAM Press, 1993, Philadelphia. H. Szu, Y. Sheng and J. Chen, "Wavelet Transform as a Bank of Matched Filters," Applied Optics,June 1992, vol. 31 no. 17, p.p. 3267-3277.
[29] J.T. Chien, "Discriminant Waveletfaces and Nearest Feature Classifiers for Face Recognition", PAMI.
[30] , 2002.
[31] Hybrid wavelet artificial neural network model for municipal water demand forecasting, Jowhar Rasheed H.M. Ibrahim,2012
[32] A new wavelet conjunction approach for estimation of relative humidity: wavelet principal component analysis combined with ANN,MaryamBayatvarkeshi, KouroshMohammadi, Ozgur Kisi &RojinFasihi, 2020
[33] Dimension Reduction of Machine Learning-Based Forecasting Models Employing Principal Component Analysis, YinghuiMeng, Sultan Noman Qasem, ManouchehrShokri and Shahab ,Mathematics,MDPI,Published: 27 July 2020

[34] Deep Neural Network for Human Face Recognition, Dr. PriyaGuptaa,NidhiSaxena,MeetikaSharma,JagritiTripathia, I.J. Engineering and Manufacturing, 2018,63-71,Published Online January 2018 in MECS, January 2018

[35] A comparative analysis of artificial neural networks and wavelet hybrid approaches to long-term toxicheavy metal prediction Peifeng Li, Pei Hua, DongweiGui, JieNiu, Peng Pei, Jin Zhang & Peter Krebs,Scientific Reports volume 10, Article number: 13439 (2020)

[36] A wavelet neural network approach to predict daily river discharge using meteorological data, ÖmerGürsoy, SerefNaciEnginFirst,SAGEJournals,April 25, 2019

[37] Graph Wavelet Neural Network,Bingbing Xu, Huawei Shen, Qi Cao, YunqiQiu, XueqiCheng,Cornell University Press,April 2019

[38] Deep Adaptive Wavelet Network,Maria Ximena Bastidas Rodriguez; Adrien Gruson; Luisa F. Polanía; Shin Fujieda; Flavio Prieto Ortiz; Kohei Takayama,2020 IEEE Winter Conference on Applications of Computer Vision (WACV),March 2020

Predictive Maintenance of Gas Turbine using Prognosis Approach

Hima Soni[1], Akshay Kansara[2], Tanvi Joshi[3]

[1]Student of 4th Semester ME at S.P.B. Patel Engineering College (Saffrony), Ahmedabad-Mehsana Highway, Gujarat, India
[2]Head of the Department at S.P.B. Patel Engineering College (Saffrony), Ahmedabad-Mehsana Highway, Gujarat, India
[3]Assistant Professor at S.P.B. Patel Engineering College (Saffrony), Ahmedabad-Mehsana Highway, Gujarat, India

---

Abstract - *Industry 4.0 revolution aimed at transforming factories from automated to smart and intelligent. To fulfil that challenge, the research idea proposed herewith, implementing prognosis approach for Predictive Maintenance on Gas Turbine. Gas turbine (GT) based turbofan engines are recognized for their high availability and reliability and are used for aero, marine and power generation applications. Maintenance of this complex machinery should be done proactively to prevent premature failure, reduce the overall cost by avoiding unnecessary maintenance task. This goal is achieved by estimating the Remaining Useful Life (RUL) of GT. The RUL is very important information to decision-makers and planners for upcoming maintenance activity. This paper aims to explore the use of neural network models to predict RUL. In recent years researchers have proposed several machine learning, data driven and neural network approaches for predicting RUL. This paper investigates the effect of the Convolutional Neural Network (CNN) in RUL Estimation. The experimental study compares this approach to purely LSTM. This result suggests the CNN is a promising model in estimating RUL of time series dataset for GT even in the case of rare events.*

Key Words: remaining useful life, prognosis approach, gas turbine, predictive maintenance

1. INTRODUCTION

1.1 Gas Turbine

Gas turbine is the combustion engine. It is working by sucking air into the front of the engine using a fan. From there, the engine compresses the air, mixes fuel with it, ignites the fuel/air mixture, and shoots it out the back of the engine, creating thrust in case of turbofan engine [2] whereas in case of powerplant gas turbine it will generate the electrical energy. Two main applications of Gas Turbine engine are Turbofan engines and powerplant gas. Gas turbine-based Turbofan engine has been considered for the present research work. Gas turbine-based turbofan is one of the complex types of machinery, running 24/7 requires effective maintenance strategy to reduce downtime and for increased reliability and availability.

1.2 Types of Maintenance

The failure of GT engine is often a significant cause of major accidents and causalities. To prevent these causes, detecting primary degradation is essential. In the field of aircraft maintenance, traditional maintenance is either purely reactive (fixing or replacing an aircraft engine component after it's complete failure) or blindly proactive (assuming a certain level of performance degradation with no input from the aircraft engine itself and maintaining the aircraft engine on a routine schedule whether maintenance is needed or not). Both scenarios are quite wasteful and inefficient, and neither is conducted in real-time [7]. These kinds of Improper maintenance may lead to an increased rate of deterioration. Due to that maintenance cost may also reach up to 35% of operating cost [5]. Scheduling of maintenance activity based on fault diagnosis, performance degradation assessment and the predicted remaining useful life of the gas turbine and the need to prevent faults in advance, prognostics and health management (PHM) is gradually replacing these two maintenance strategies. [7]

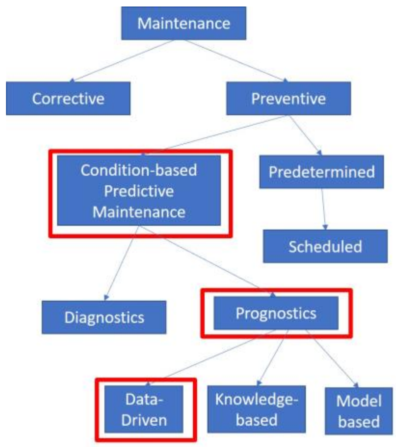


Fig-1: Types of Maintenance

1.3 Predictive Maintenance using IoT

Simple Predictive maintenance is a maintenance methodology that involves monitoring the health of a machine and applying predictive modelling techniques in order to predict the likelihood of failure of the machine and a time estimate of its probable occurrence [6]. Whereas, Predictive maintenance in the Internet of Things (IoT) era can be summarized as a maintenance methodology that brings together the power of machine learning and streaming sensor data to maintain machines before, they fail, optimize resources, and thereby reduce unplanned downtime. This is where IIoT helps in terms of providing all the data in the framework to make a meaningful assessment. [6]

1.4 Prognosis Approach

The main reason of sudden failures is that the diagnostics system cannot catch fault progression. To avoid such failures, the maintenance strategy needs to be changed from fail and fix to predict and prevent. In other words, instead of being reactive to be proactive. Prognostics is the main driver to proactivity [8]. Prognostics can be defined as a remaining useful life (RUL) estimation process of system/subsystem/component. In the beginning of life (BoL), the system operates normally with its full health. When the system starts to degrade, this can be considered as a trigger point to the prognostics system. The prognostics system continues to operate and performs RUL estimation at subsequent prediction points. In the same time, prognostics information is the main driver for condition-based maintenance (CBM) [8].

1.5 Estimating Remaining Useful Life (RUL)

There are three main classes of RUL prediction methods: (1.) data-driven methods, (2.) physics model-based methods, and (3.) methods that combine data-driven and physics model-based methods. The data-driven methods use past condition monitoring data, the current health status of the system, and data on the degradation of similar systems. There are two main challenges in prognostics based on physics: (1.) there is not enough physical knowledge to construct a physical degradation model and (2.) the values of the physical model's parameters are difficult to determine exactly. Therefore, it is important to understand the failure mechanism of the system correctly, and experienced personnel are required for physics-based models. Therefore, the requirements of data-driven methods to model the degradation and predict the RUL are easier to satisfy. [7]

The performance of many data-driven prognostics methods is heavily dependent on the choice of the performance degradation data to which they are applied. However, engines have many sensor parameters. The sensitivity of the data from different sensors varies in terms of showing engine performance degradation; the data from some sensors are sensitive and the data from other sensors are not sensitive. Therefore, it is necessary to select suitable sensor parameters whose data are more sensitive to the engine's performance degradation trend as the training data for the RUL prediction model. [7]

Three problems hinder the implementation of performance degradation feature extraction in practice. The traditional methods of extracting performance degradation features for prognostics are unsupervised and cannot automatically adjust the feature extraction modal parameters based on feedback from the prediction. Such feature extraction and choice are significant but represent a principal shortcoming of popular prognostics algorithms: the inability to extract and organize discriminatively or trend information from data. Therefore, it is important to develop an automatic feature extraction method that can extract the prominent feature to achieve better insight into the underlying performance degradation state. Deep learning, a new method that has been put forward in the last few years, can be used to extract multilevel features from data, which means the method could express data at different levels of abstraction. Deep learning is an end-to-end machine learning system.[7] Based on the present literature survey Convolutional Neural Network (CNN) is the one which is selected among all the Deep neural network approaches for the present research. CNN accepts the image as an input for that reason time series to image conversion is to be done. There are so many methods available to transform time series to images.

1.6 Imaging Timeseries

Recently, great results have been achieved by processing data with deep learning techniques, and, specifically, by using convolutional neural networks (CNN) with images as input. In scenarios where input data isn't formatted as an image, many transformation methods have helped apply CNNs to other data types. Time series is one of these data structures that can be modelled to approach the problem from a computer vision perspective.[28]

Recurrence plots are an advanced technique for visually representing multivariate non-linear data. This refers to a graph representing a matrix, where elements correspond to those times at which the data recurs to a certain state or phase. Recurrent behavior, such as periodicities or irregular cyclicities, is a fundamental property of deterministic dynamical systems, like non-linear or chaotic systems. As higher dimensional datasets can't be pictured easily, they can only be visualized by projection onto 2D or 3D sub-spaces. Recurrence plots enables the visualization of the mm-dimensional phase space through a two-dimensional representation of its recurrence. This recurrence of a certain state at time ii at a different time jj is marked within a 2D squared matrix and can be mathematically expressed as: [28]

$$R_{i,j} = \theta\,(\varepsilon_i - ||\,\vec{x_i} - \vec{x_j}\,||), \vec{x_i} \in R^m, \qquad i,j = 1, \ldots, N$$

The main advantage of using recurrence plots is being able to visually inspect any higher dimensional phase space trajectories by obtaining an image that hints at how the series evolve over time. [28]

2. PRIOR ART SEARCH

While proceeding with the research work, various approaches for the Predictive Maintenance using Prognosis approach have been analyzed in the prior art search. For predictive maintenance on Gas Turbine total, 4 research papers have been referred. As shown earlier predictive maintenance can be divided further in two parts. One is a diagnosis and other is the prognosis. Now prognosis can be done using Machine Learning approach, Data driven approach or even using neural network. In the below table all the literatures are classified in different approaches.

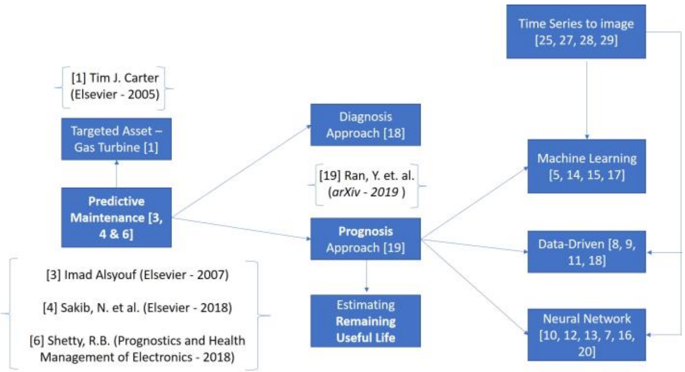


Fig-2: Taxonomy of Prior Art Search

Machine Learning	Data Driven	Neural Network
Autoregressive Model *Ahsan, S. et. al. (EDP Sciences - 2017) [8]*	**Data Driven Prognosis** *Elattar, H.M. et. al. (Springer - 2018) [11]*	**Directed Acyclic Graph Network Combined With CNN And LSTM** *Li, J. et. al. (IEEEAccess - 2019) [20]*
Bayesian Approaches *Mosallam, A. (Doctoral dissertation - 2014)[14]*	**Performance Based Health Monitoring** *Tahan, M. et. al. (Elsevier - 2017)[9]*	**Semi-supervised Deep Architecture** *Ellefsen, A.L. et. al. (Elsevier - 2019) [12]*
Data-level and Feature-level Fusion Ghorbani, S. et. al. (Springer - 2020) [17]	**Data-based Investigation On The Performance Of An Independent Gas Turbine** *Esan, A.B. et. al. (Elsevier - 2019) [8]*	**Multilayer Perceptron & Kalman Filter** *Alberto-Olivares, M. et. al. (IEEE - 2019) [13]* **Stacked Sparse Autoencoder With Multilayer Self-learning** *Ma, J. et. al. (Hindawi-Complexity - 2018) [7]*
Similarity Model & Support Vector Machine Approach *Chen, Z. et. al. (Energies 2018) [15]*	**Ensemble Of Data-driven Prognostic Algorithms** *Hu, C. et. al. (Elsevier - 2012) [18]*	**Hybrid Deep Neural Network** *Al-Dulaimi, A. et. al. (IEEE - 2019) [16]* **Transfer Learning** *Fan, Y. et. al. (arXiv - 2019) [10]*

Fig-3: Classification of Literatures in different Approaches for Prognostics

The purpose of this paper is to investigate the use of Deep learning-based models, to find remaining operational cycles (RUL) before failure in the test set in case of GT, which is the main objective of the present research work. As a supportive objectives Prognosis approach is used for performing predictive maintenance, Recurrence Plot method to transfer time series into images and Convolutional Neural Network has been selected to classify the data to find the remaining operational cycle.

3. DATASET DESCRIPTION

The Turbofan Engine Degradation Simulation Data Set by the Prognostic Center of Excellence of National Aeronautics and Space Administration (NASA) is used in the present research work. This data set was created by synthetic data collected from a thermodynamic simulation model called C-MAPSS (Commercial Modular Aero-Propulsion System Simulation). The simulator consists of 14 input parameters and 21 output parameters are reported in the data set. Each turbofan unit provides the following information: [13] The data set consists of multiple time series divided into 4 training and test subsets, both identified by the names: FD001, FD002, FD003 and FD004 (see fig-4). The whole data set contains sensor data from several turbofans, which operate normally at the beginning of the recording and eventually they develop a failure. [13].

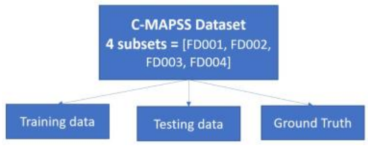


Fig-4 (a)

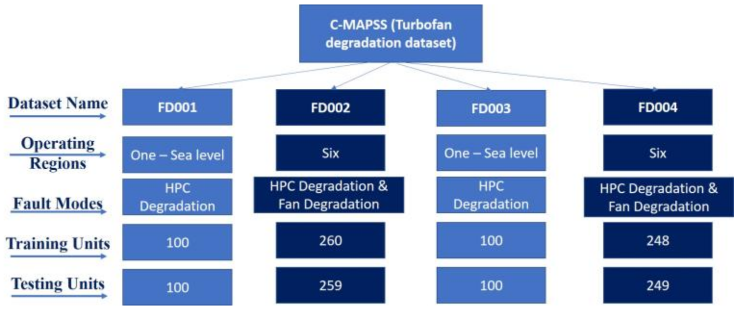


Fig-4 (b)

Fig-4: Dataset Description

There are two failure modes: high-pressure compressor degradation and fan degradation. As per the fig-4 (a).

4. METHODOLOGY

This section introduces the relevant procedure used in this research. As shown in Fig -5, the whole procedure for RUL prediction for a gas turbine-based turbofan engine consists of two main steps: data pre-processing and RUL prediction using CNN & LSTM.

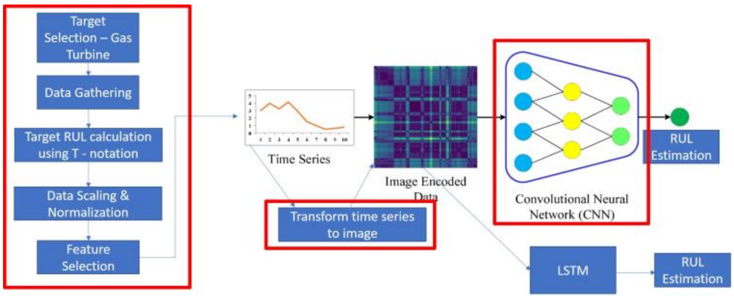


Fig-5: Overall Framework

Once the target asset selection is done, dataset must get prepared before applying Machine Learning Models. Data gathering, RUL calculation, data scaling and normalization and feature selection all these steps come under data preparation/preprocessing part. As CNN is only taking images as an input, so in this present research work timeseries data has been converted into the images. This will be one additional preprocessing step in case of CNN. At last training, testing and evaluation of the model will be held. Finally, RUL estimation shown as an output.

4.1 Data Gathering

Usually, the gas turbine-based turbofan engine is operating normally, at the starting of each time series and develops a fault at some point of time during the series. In the training set, the data recording ends when the turbofan stops working completely due to the failure, representing a data base of turbofans that failed during operation. On the other hand, in the test set, the time series terminates sometime before the system stops working, representing a data base of turbofans in the current operation.

4.2 Data Preparation

Feature Selection, Data Normalization & RUL calculation are commonly coming under the Data Preparation step. Training & testing dataset will be pre-processed to achieve better results having good accuracy.

4.2.1 Feature Selection

Different sensors in gas turbine-based turbofan engine are having very different responses to the performance degradation process. Some sensors show unclear inclinations because of noise or insensitivity to degradation trends. Choosing unresponsive parameter data may reduce the RUL estimation accuracy. To improve the performance of the estimation model, sensors that are more responsive to the performance degradation process are chosen as inputs to the RUL estimation model.

4.2.2 Prepare Target columns

To train the model for RUL estimation, it is necessary to have a set of input and output data columns. Where the input data is the information recorded from several sensors and the output data is the RUL. However, databases for prognosis applications do not often contain the RUL information for training because, in many industrial applications, it is impossible to accurately assess the RUL information. Therefore, one should derive the RUL column (i.e. the time remaining before the end of each turbofan data recording), it can be derived as follows:

RUL Derivation:

$$LastCycle(Unitnumber) - CurrentCycle(i) = RemainingUsefulLife(RUL)$$

$$X - u_i = RUL, i = 1, \ldots, n$$

Let say, 'u' is unit number and each unit contain 'i' number of cycles. To calculate RUL, firstly the last cycle of each unit has been searched which denoted here as X. Likewise, RUL for each row will be calculated. In the case of regression, RUL column having continuous values is enough.

Labelling:

In the case of classification approach, one more column containing class labels needs to be derived. Which will be derived from RUL column.

For 0 to 15 remaining cycles, the given label is 2, 16 to 45 remaining cycles are labelled as 1 and for the RUL which are greater than 45 will be classified as 0. It is clear that in reality, the category labelled as 2 is the most economically valuable. Except for the convolutional neural network, data preparation part for other models are done up to this step.

4.2.3 Data Normalization:

After the elimination of some constant columns and selection of informative columns, the linear function (i.e. min-max normalization function) that best preserves the original performance degradation pattern of the aircraft engine is chosen to map the data for each selected sensor to [-1, 1].

4.3 Transform Time series to images

4.3.1 Recurrence Plot

As CNN model is being used to estimate the RUL, the transformation of whole preprocessed time series dataset into images are must be needed. That has been achieved with the help of the recurrence plot method.

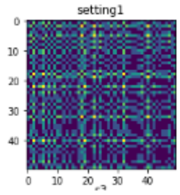


Fig-6: Sample Recurrence Plot of size 50x50x17

This image is an example of a recurrence plot which is nothing but the resultant representation.

To achieve the recurrence plot first extraction of the time series sequence is to be done.

4.3.2 Generating Sequence for transforming time-series into images

As an example, Unit 1 contains 192 cycles and consider window size as a 50. As a result of this generating sequence process, the total 142 sequences are generated of 50 window size as shown in fig-7.

Fig-7: Scenario of Timeseries Window Generation

Finally, at this stage preprocessing part for every Machine Learning model is completed.

4.4 Model Selection & Building

4.4.1 Model 1: CNN

Convolutional neural networks (CNNs) are a category of neural networks that have proven very effective in areas such as image recognition and classification. CNNs have been successful in identifying faces, objects, and traffic signs in addition to powering vision in robots and self-driving cars. CNNs derive their name from the "convolution" operator. The primary purpose of convolution in the case of CNNs is to extract features from the input image. the model learns how to automatically extract the features from the raw data that are directly useful for the problem being addressed. This is called "representation learning". The ability of CNNs to learn and automatically extract features from raw input data can be applied to time series forecasting problems. A sequence of observations can be treated like a one-dimensional image that a CNN model can read and refine into the most appropriate elements. This capacity of CNN has been proved to great effect on the time series classification task of turbo fan's remaining operational cycles prediction [26].

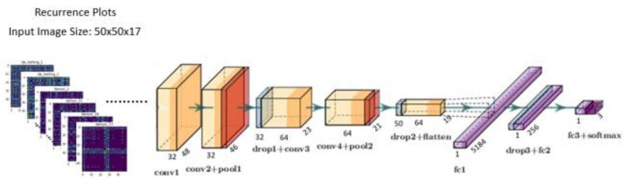


Fig-8: CNN Model Architecture

4.4.2 Model 2: LSTM

The Long Short-Term Memory network (LSTM) network, is a recurrent neural network that is trained using Backpropagation Through Time and overcomes the vanishing gradient problem. Instead of neurons, LSTM networks have memory blocks that are connected through layers. LSTM is capable of learning long-term dependencies. LSTMs are explicitly designed to avoid the long-term dependency problem. Remembering information for long periods is practically their default behavior, not something they struggle to learn! All recurrent neural networks have the form of a chain of repeating modules of the neural network. In standard RNNs, this repeating module will have a very simple structure, such as a single tanh layer [35].

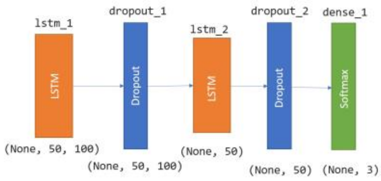


Fig-9: LSTM Model Architecture

4.5 Training, testing & validation

In this phase, by considering the randomly initialized parameters, training will be done. While training, cross-validation is done to avoid overfitting. Once the model is trained, parameters will get optimized for the next training and model will be retrained with the best parameters. Finally, the evaluation of the model will be held on the testing dataset.

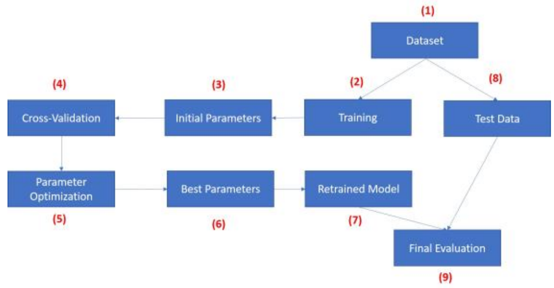


Fig-10: Flow of Training, Testing & Evaluation

4.6 RUL estimation

Finally, RUL will be estimated like, how much time (in terms of the number of cycles) is left before the next fault?

5. EXPERIMENTAL STUDY, RESULTS & DISCUSSION

All experiments are operated on Google Collaboratory platform having GPU. Using Colab individual can import an image dataset, train an image classifier upon it, and evaluate the model. Colab notebooks execute code on Google's cloud servers, that means one can leverage the power of Google hardware, including GPUs and TPUs. Also, Keras libraries are used having TensorFlow backend. Keras is a powerful open-source Python library for developing and evaluating deep learning models such as CNN and LSTM. It covers the efficient numerical computation libraries Theano and TensorFlow used for machine learning applications such as neural networks.

5.1 Understanding of the dataset through the plots

Once the CMAPSS dataset column labelling is done, training, testing and ground truth data sets are loaded. Data are available in the form of time series. Here in the below plot, one can conclude that unit 69 has the maximum number of cycles and unit 39 has the minimum number of cycles.

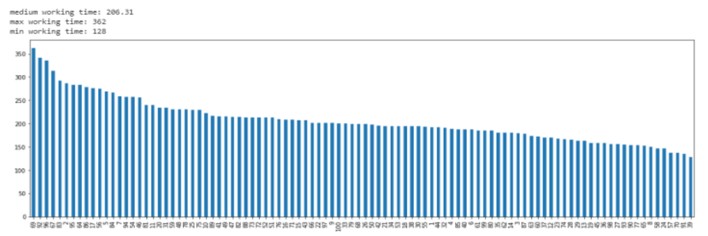


Fig-11: Plotting for the number of cycles of each unit

Just by plotting, identification of some data attribute containing constant data values can be done. Based on the identification, all the constant features have been eliminated in this step. To plot is always a good idea, in this way, one can have an impressive and general overview of the data at our disposal.

Fig-12: Plotting of Dataset Attributes

Verified feature selection with Correlation Matrix for all engine units. Out of all the columns these 17 columns are selected for further analysis: 'Op_Setting_1', 'Op_Setting_2', 'Sensor_2', 'Sensor_3', 'Sensor_4', 'Sensor_6', 'Sensor_7', 'Sensor_8', 'Sensor_9', 'Sensor_11', 'Sensor_12', 'Sensor_13', 'Sensor_14', 'Sensor_15', 'Sensor_17', 'Sensor_20', 'Sensor_21'

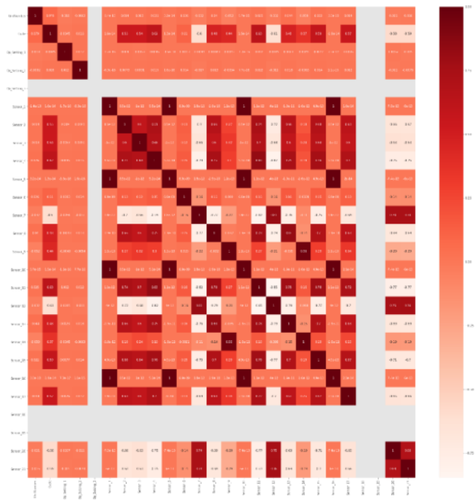


Fig-13: Correlation Matrix for Feature Selection

5.2 RUL Calculation of each row

This section calculates RUL in T-Minus notation. As covered in the methodology section, by finding the last cycle of each unit and then subtracting of a current cycle from it will results in RUL for each row. In case of training set, last cycle (time of failure) is provided in the data set itself.

	UnitNumber	Cycle	Op_Setting_1	Op_Setting_2	Op_Setting_3	Sensor_1	Sensor_2	....	Sensor_21	Target_Remaining_Useful_Life
0	1	1	-0.0007	-0.0004	100.0	518.67	641.82		23.4190	191
1	1	2	0.0019	-0.0003	100.0	518.67	642.15		23.4236	190
2	1	3	-0.0043	0.0003	100.0	518.67	642.35		23.3442	189
3	1	4	0.0007	0.0000	100.0	518.67	642.35		23.3739	188
4	1	5	-0.0019	-0.0002	100.0	518.67	642.37		23.4044	187
...	...	...	...	...	...	...	...		...	...
20626	100	196	-0.0004	-0.0003	100.0	518.67	643.49		22.9735	4
20627	100	197	-0.0016	-0.0005	100.0	518.67	643.54		23.1594	3
20628	100	198	0.0004	0.0000	100.0	518.67	643.42		22.9333	2
20629	100	199	-0.0011	0.0003	100.0	518.67	643.23		23.0640	1
20630	100	200	-0.0032	-0.0005	100.0	518.67	643.85		23.0522	0

20631 rows × 27 columns

Fig-14: Snapshot of the training dataset

In case of test set the last cycle of each unit (i.e. time to failure(ttf)) is not given in the test set. So, to calculate the RUL of each raw, values given in the ground truth set will be considered as the value of the last cycle (ttf).

	UnitNumber	Cycle	Op_Setting_1	Op_Setting_2	Sensor_2	Sensor_3	....	Sensor_21	Target_Remaining_Useful_Life
0	1	1	0.65625	0.692308	0.596215	0.421968		0.620099	142
1	1	2	0.34375	0.230769	0.182965	0.504025		0.645718	141
2	1	3	0.53125	0.538462	0.419558	0.464814		0.681104	140
3	1	4	0.77500	0.461538	0.413249	0.391587		0.620382	139
4	1	5	0.60000	0.461538	0.435331	0.471306		0.676008	138
...	...	...	...	...	...	...		...	...
13091	100	194	0.81875	0.461538	0.665615	0.789665		0.370842	24
13092	100	195	0.44375	0.384615	0.659306	0.692028		0.483652	23
13093	100	196	0.47500	0.230769	0.728707	0.626071		0.381741	22
13094	100	197	0.27500	0.538462	0.671924	0.673851		0.473461	21
13095	100	198	0.59375	0.692308	0.574132	0.846014		0.353999	20

13096 rows × 22 columns

Fig-15: Snapshot of the testing dataset

5.3 Adding labels to the dataset

The class label has been given based on RUL column containing continuous values, as follows.

	UnitNumber	Cycle	Op_Setting_1	Op_Setting_2	Op_Setting_3	Sensor_1	....	Target_Remaining_Useful_Life	label2
0	1	1	-0.0007	-0.0004	100.0	518.67		191	0
1	1	2	0.0019	-0.0003	100.0	518.67		190	0
2	1	3	-0.0043	0.0003	100.0	518.67		189	0
3	1	4	0.0007	0.0000	100.0	518.67		188	0
4	1	5	-0.0019	-0.0002	100.0	518.67		187	0
...	...	...	...	...	...	...		...	...
20626	100	196	-0.0004	-0.0003	100.0	518.67		4	2
20627	100	197	-0.0016	-0.0005	100.0	518.67		3	2
20628	100	198	0.0004	0.0000	100.0	518.67		2	2
20629	100	199	-0.0011	0.0003	100.0	518.67		1	2
20630	100	200	-0.0032	-0.0005	100.0	518.67		0	2

20631 rows × 29 columns

Fig-16: Snapshot of Dataset after adding label column

5.4 Data Normalization

This step of data preparation is made using min-max pooling as discussed in the methodology section. Min-max normalization has been used to enable the unbiased contribution from the output of each sensor, i.e.,

$$xi = \frac{2(x_i - min x_i)}{max\, x_i - x_i} - 1,$$

where x_i is the time sequence of ith sensor measurements, and xi is the normalized sensor data. This normalization will guarantee equal contribution from all features across all operating conditions [34]. The normalized data will be between [-1,1].

5.5 Time Series to image transformation

gen_sequence() and gen_labels() functions are used to generate sequences from time series and label each window of size 50 respectively as shown in the below code snippet. For efficiency reason a 2D CNN requires spatial invariance. So, to transform the time series windows to images Recurrence Plots has been used. They are easy to implement in python Scipy with a few lines of code.

```
sequence_length = 50

def gen_sequence(id_df, seq_length, seq_cols):

    data_matrix = id_df[seq_cols].values
    num_elements = data_matrix.shape[0]
    for start, stop in zip(range(0, num_elements-
seq_length), range(seq_length, num_elements)):
        yield data_matrix[start:stop, :]

def gen_labels(id_df, seq_length, label):

    data_matrix = id_df[label].values

   num_elements = data_matrix.shape[0]
    return data_matrix[seq_length:num_elements, :]
```

5.6 Recurrence Plot

Using that function on time series sequence images are generated of size 50x50. One observation is made by an array of images of size 50x50x17. Where 17 is the number of non-zero variance columns.

```
def rec_plot(s, eps=0.10, steps=10):
    d = pdist(s[:,None])
    d = np.floor(d/eps)
    d[d>steps] = steps
    Z = squareform(d)
    return Z
```

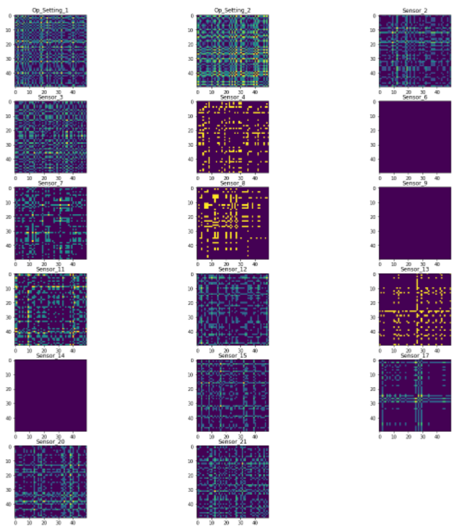


Fig -17: Snapshot of some observations

5.7 Model fitting & evaluation

5.7.1 CNN

Model is trained using 9 epochs using early stopping callbacks and to avoid overfitting cross-validation is applied. Adam optimizer is being used to tune and optimize the hyperparameter. Also, cross-validation is used to avoid overfitting. Finally using early stopping callbacks model get trained at 9th epoch only. Once model gets fitted now it's time to evaluate the model using a test set. While evaluating the trained model in case of FD001 dataset 91.74% accuracy is achieved.

5.7.2 LSTM:

Model is trained using 3 epochs using early stopping callbacks and to avoid overfitting cross-validation is applied. Once model gets fitted now it's time to evaluate the model using a test set. While evaluating the trained model in case of FD001 dataset 79.91% accuracy is achieved.

5.8 Results

In this sub-section, various experimental results have been presented to evaluate the performance of the CNN model for RUL estimation. Also, following the same steps, the model for FD002, FD003 & FD004 datasets has been trained and for every dataset, accuracy have been compared with LSTM model's resultant accuracy in the next subsection (5.9).

Table-1: Accuracy & Loss with CNN Model

Dataset Name	Accuracy	Loss
FD001	91.74%	23.28%
FD002	87.74%	36.33%
FD003	94.16%	19.24%
FD004	93.64%	39.50%

5.9 Comparison between CNN & LSTM

Table-2: Accuracy Comparison between CNN & LSTM

Accuracy comparisons Table				
Approaches used	**FD001**	**FD002**	**FD003**	**FD004**
CNN	91.74%	87.74%	94.16%	93.64%
LSTM	79.91%	80.50%	90.24%	74.18%

Implementing both models on the same preprocessed time-series data, it is found that CNN model is giving better accuracy than LSTM at the same time research objective has been fulfilled of classifying very rare events which are challenging to do with LSTM. CNNs are also good for feature extraction for the same reason, making them beneficial for transfer learning. For the system on which rarer events prediction is require in that way the decision has been made to go with CNN.

6. CONCLUSION AND FUTURE WORK

In this study, deep CNN and recurrence plot for the gas turbine-based turbofan engine have been explored. Segmentation of the time series dataset has been performed and generated recurrence plot image to train a deep CNN.

The system achieved accuracy in the range of 91.74% to 94.16% in benchmark dataset. Observation has been made that a deep CNN can learn recurrence plot from historical sensor data and can make a remaining useful life estimation. Using which one can achieve the following benefits,

1. Feature extraction is automatic.
2. In case of any prognostic health management dataset, one can train the applied CNN model with the use of user-friendly GUI and estimate RUL of the targeted assets, due to automatic feature extraction. Even though the user does not have much knowledge about the dataset and ML.
3. It will reduce the downtime as well as money loss. At the same time, it will increase the efficiency of the turbofan engine with the Just in Time maintenance.
4. Using Convolutional Neural Network one can classify the rare events using time-series data.

Overall, this work has demonstrated the performance of deep CNN to learn a recurrence plot pattern and estimate RUL. The future work will focus on improving the accuracy of the model. Besides, one can explore the performance to real-time data.

REFERENCES

[1] Rani, S., 2018. Common failures in gas turbine blade: A critical review. IJESRT (International Journal of Engineering Sciences and Research Technology, vol. 7, no. 3, pp. 799-803, 2018. doi: 10.5281/zenodo. 1207072.

[2] Boldmethod.com. (2019). How Does A Turbofan Engine Work? [online] Available at: https://www.boldmethod.com/learn-to-fly/aircraft-systems/how-does-a-jet-engine-turbofan-system-work-the-basics/.

[3] Alsyouf, I., 2007. The role of maintenance in improving companies' productivity and profitability. International Journal of production economics, 105(1), pp.70-78.

[4] Sakib, N. and Wuest, T., 2018. Challenges and Opportunities of Condition-based Predictive Maintenance: A Review. Procedia CIRP, 78, pp.267-272.

[5] Ahsan, S. and Lemma, T.A., 2017. Remaining Useful Life Prediction of Gas Turbine Engine using Autoregressive Model. In MATEC Web of Conferences (Vol. 131, p. 04014). EDP Sciences.

[6] Shetty, R.B., 2018. Predictive Maintenance in the IoT Era. Prognostics and Health Management of Electronics: Fundamentals, Machine Learning, and the Internet of Things, pp.589-612.

[7] Ma, J., Su, H., Zhao, W.L. and Liu, B., 2018. Predicting the remaining useful life of an aircraft engine using a stacked sparse autoencoder with multilayer self-learning. Complexity, 2018.

[8] Esan, A.B., Ehiaguina, V., Awosope, C., Olatomiwa, L. and Egbune, D., 2019. Data-based investigation on the performance of an independent gas turbine for electricity generation using real power measurements and other closely related parameters. Data in brief, 26, p.104444.

[9] Tahan, M., Tsoutsanis, E., Muhammad, M. and Karim, Z.A., 2017. Performance-based health monitoring, diagnostics and prognostics for condition-based maintenance of gas turbines: A review. Applied energy, 198, pp.122-144.

[10] Fan, Y., Nowaczyk, S. and Rögnvaldsson, T., 2019. Transfer learning for remaining useful life prediction based on consensus self-organizing models. arXiv preprint arXiv:1909.07053.

[11] Elattar, H.M., Elminir, H.K. and Riad, A.M., 2018. Towards online data-driven prognostics system. Complex & Intelligent Systems, 4(4), pp.271-282.

[12] Ellefsen, A.L., Bjørlykhaug, E., Æsøy, V., Ushakov, S. and Zhang, H., 2019. Remaining useful life predictions for turbofan engine degradation using semi-supervised deep architecture. Reliability Engineering & System Safety, 183, pp.240-251.

[13] Alberto-Olivares, M., Gonzalez-Gutierrez, A., Tovar-Arriaga, S. and Gorrostieta-Hurtado, E., 2019, September. Remaining Useful Life Prediction for Turbofan based on a Multilayer Perceptron and Kalman Filter. In 2019 16th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE) (pp. 1-6). IEEE.

[14] Mosallam, A., 2014. Remaining useful life estimation of critical components based on Bayesian Approaches (Doctoral dissertation).

[15] Chen, Z., Cao, S. and Mao, Z., 2018. Remaining useful life estimation of aircraft engines using a modified similarity and supporting vector machine (SVM) approach. Energies, 11(1), p.28.

[16] Al-Dulaimi, A., Zabihi, S., Asif, A. and Mohammadi, A., 2019, May. Hybrid Deep Neural Network Model for Remaining Useful Life Estimation. In ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 3872-3876). IEEE.

[17] Ghorbani, S. and Salahshoor, K., 2020. Estimating Remaining Useful Life of Turbofan Engine Using Data-Level Fusion and Feature-Level Fusion. Journal of Failure Analysis and Prevention, 20(1), pp.323-332.

[18] Hu, C., Youn, B.D., Wang, P. and Yoon, J.T., 2012. Ensemble of data-driven prognostic algorithms for robust prediction of remaining useful life. Reliability Engineering & System Safety, 103, pp.120-135.

[19] Ran, Y., Zhou, X., Lin, P., Wen, Y. and Deng, R., 2019. A Survey of Predictive Maintenance: Systems, Purposes and Approaches. arXiv preprint arXiv:1912.07383.

[20] Li, J., Li, X. and He, D., 2019. A directed acyclic graph network combined with cnn and lstm for remaining useful life prediction. IEEE Access, 7, pp.75464-75475.
[21] Saxena, A., Goebel, K., Simon, D. and Eklund, N., 2008, October. Damage propagation modeling for aircraft engine run-to-failure simulation. In 2008 international conference on prognostics and health management (pp. 1-9). IEEE.
[22] Yang, C.L., Chen, Z.X. and Yang, C.Y., 2020. Sensor Classification Using Convolutional Neural Network by Encoding Multivariate Time Series as Two-Dimensional Colored Images. Sensors, 20(1), p.168.
[23] Hailesilassie, T., 2019. Financial Market Prediction Using Recurrence Plot and Convolutional Neural Network.
[24] Li, X., Ding, Q. and Sun, J.Q., 2018. Remaining useful life estimation in prognostics using deep convolution neural networks. Reliability Engineering & System Safety, 172, pp.1-11.
[25] Estebsari, A. and Rajabi, R., 2020. Single Residential Load Forecasting Using Deep Learning and Image Encoding Techniques. Electronics, 9(1), p.68.
[26] Lazzeri, F. (2019). 3 reasons to add deep learning to your time series toolkit. [online] O'Reilly Media. Available at: https://www.oreilly.com/content/3-reasons-to-add-deep-learning-to-your-time-series-toolkit/ [Accessed 27 May 2020].
[27] pyts.readthedocs.io. (n.d.). 4. Imaging time series — pyts 0.12.dev0 documentation. [online] Available at: https://pyts.readthedocs.io/en/latest/modules/image.html [Accessed 27 May 2020].
[28] Datascience.aero. (2019). From time to space! Time series converted into images to train CNNs. [online] Available at: https://datascience.aero/time-space-time-series-train-cnn/ [Accessed 27 May 2020].
[29] kaggle.com. (n.d.). RecuPlots and CNNs for time-series classification. [online] Available at: https://www.kaggle.com/tigurius/recuplots-and-cnns-for-time-series-classification [Accessed 27 May 2020].
[30] kaggle.com. (n.d.). Predictive Maintenance using LSTM on sensor data. [online] Available at: https://www.kaggle.com/nafisur/predictive-maintenance-using-lstm-on-sensor-data [Accessed 27 May 2020].
[31] Alhamaly, A. (2019). Jet Engine Remaining Useful Life (RUL) Prediction. [online] Medium. Available at: https://medium.com/@hamalyas_/jet-engine-remaining-useful-life-rul-prediction-a8989d52f194 [Accessed 27 May 2020].
[32] yajasd.github.io. (n.d.). Predicting Engine Failure using C-MAPSS data · Yajas Dwivedi. [online] Available at: https://yajasd.github.io/2018/06/04/Predicting-Engine-Failure/ [Accessed 27 May 2020].
[33] Cerliani, M. (2020). Remaining Life Estimation with Keras. [online] Medium. Available at: https://towardsdatascience.com/remaining-life-estimation-with-keras-2334514f9c61 [Accessed 27 May 2020].
[34] Ksenia Sorokina (2017). Image Classification with Convolutional Neural Networks. [online] Medium. Available at: https://medium.com/@ksusorokina/image-classification-with-convolutional-neural-networks-496815db12a8 [Accessed 15 Jun. 2019].
[35] Github.io. (2015). Understanding LSTM Networks -- colah's blog. [online] Available at: https://colah.github.io/posts/2015-08-Understanding-LSTMs/.