International Conference on Recent Trends in Data Science and its Applications
DOI: rp-9788770040723.178

Visualizing Teams Performance in National Kabaddi League through Dimensionality Reduction Techniques

M.Gayathri[1],V.Kavitha[2],

[1,2]*Department of Data Science and Business System, SRM Institute of Science and TechnologyAddress*

[1]gaya.eng@gmail.com[2]ka
vithav2@srmist.edu.in

***Abstract*—**The aim of this paper, is to gain insights in to dynamicsofKabaddi.Thissporthasbecomeincreasinglypopular, by analyzing teams' performance in national leagues. To do so, theresults of each round of matches the participating teams play are usedto create a dissimilarity matrix. This matrix is then processed throughtwo algorithms, Multidimensional Scaling (MDS) and t-DistributedStochasticNeighborEmbedding(t-SNE),tovisualizetheperformanceofeachteam.TheMDSalgorithmprovidesalow-dimensional representation of the teams' performance, allowing for avisual representation of complex data linkages. The t-SNE algorithm,on the other hand, captures the non-linear relationships in the teams'performance. By using data from the 2017-2018 Kabaddi season andcomparing the results obtained through MDS and t-SNE, this studyaims to identify critical factors that influence team performance andprovideadeeper understandingof thedynamicsofthesport.

***Keywords*— Data Visualization, Feature Projection, Sportsanalysis,Statistical analysis.**

I. INTRODUCTION

The digital data produced in various industries has grownexponentially in recent years. Teams of seven players eachcompeteinthepopularsportofKabaddionarectangularpitch. There is a 5-minute intermission in between each of the twohalves of the game's two 20-minute halves. To outscore youropponent, you must send raiders through their defense, touchas many players as you can on the way back to your side, andtackle your opponent's raiders to stop them from scoring. Theteam with the greatest number of points at the conclusion ofthe contest is deemed the winner. To make the informationcontainedinhigh-dimensionaldataeasilyaccessibleandcomprehensible, visualization is crucial. However, because thedata is computationally costly and has numerous dimensions,this can be a difficult task.[1]. Despite the challenges faced byhuge, multi-dimensional datasets, the display of data plays akey role in understanding and interpreting it. Although it canbe difficult to spot patterns and relationships because of theuncertainty in these sorts of data, visualization tools enable thediscovery of insights that might not be immediately apparent.A deeper comprehension of the facts can be attained by visualexploration, resulting in better decision-making. Additionally,data visualization can assist in identifying areas that requireadditionalresearchandanalysis.[2].Formanyyears,approaches for minimizing the dimension of the data, such asPCAandMDS,havebeenessentialtools.FeatureMappingis atechniquethataimstosimplifydatabycreatingacompressedrepresentation that retains as much of the original form aspossible. The goal is to maintain the important features andrelationshipswithinthedatawhilereducingitscomplexityandpresentingitinamoremanageableformat.Thedataarefrequentlyreducedtotwoorthreedimensionsduetothisprocedure. It is frequently visualized using scatterplots, whichplacecomparabledatasets closetogetherandunlikeonesapart. [1]. The concept of "t-SNE" is offered as a method forvisualizing high-dimensional data transformed into a matrix ofsimilarities between pairs of points. While also revealing moresignificant patterns and clusters in the data at various sizes, t-SNEisefficientincapturingthelocalizedstructureofthedata.[3]. MDS is an analytical tool that examines the variances orranges between quantitative points to help visualize complexinformation. It seeks to decrease data dimensionality whilemaintainingconsistentlinksbetweendataelements.Itmakes itsimple to visualize and analyze the data by displaying it in areduced-dimension space. Multidimensional scaling (MDS) isa technique employed in data analysis that aims to understandand interpret high-dimensional data. This approach identifiespatternsandlinkagesthatmightnotbeimmediatelyobviousinthe actual data by examining the distances between data pointsand modeling the data in a lower-dimensional space. MDS hasa wide range of applications in various fields such as scientificcomputing, computational linguistics, biostatistics, and imageprocessing. Its purpose is to simplify and present informationinamoreunderstandableandnaturalformat[4].

II. RELATEDRESEARCH

Incontrasttotypicalmultivariatedata,whichonlyconsiders individual items, multidimensional scaling (MDS) isassessing data that refers to the similarity or distinction amongpairsofdata's.[5].Multidimensionalscaling(MDS)isadimensionality reduction algorithm that analyses the similarityor difference between pairs of objects in a dataset, as opposedtotypicalmultivariatedataanalysiswhichonlylooksatindividual items.[5]. A novel method for assessing non-metricMDS was createdby Taguchi and Oono [6] and utilized tofindpatterns.ResultsusingnonparametricMDShighlydependontheinitialconfigurationbecauseitjustpreservesthesequenceofcommonalitiesratherthantheoriginalscale.

DOI: rp-9788770040723.178

Despite this, it is still difficult to shorten the computation formetrics.TheSupervisedt-StochasticNeighborEmbeddingtechnique proposed a new metric for determining dissimilaritythat takes class information into account. With the updated t-SNEmethod,high-dimensionaldatacanbeeffectivelyprocessedforvisualrepresentation,dataextraction,andclassification purposes within classification applications. [7].Thet-SNEtechniqueaimstopreservetherelationshipbetween features by transforming high-dimensional data into alower-dimensionalrepresentation.Theresultisavisualrepresentation in which points that were close together in thehigh-dimensional data will still be close together, and thosethat were far apart will remain so in the lower-dimensionalrepresentation. Maintaining the Integrity of the Specifications.Pezzottiet.Al.[8]aimedtosolvetheissueoffinding approximatenearestneighboursinlargedimensionsbyutilizing adecision tree of approximation K-d trees, whichresultsinafastercalculationprocess.Currently,therearemanyadvancedtechniquesforselectingapproximateneighbours in multi dimensions that make the computationeven moreefficient.

III. DATADESCRIPTION

The national pro kabaddi league can be found on the website"https://www.prokabaddi.com," which provides a database ofmatch results, team names, final scores, top scorers, and otherdetails of each match. The league, which is now in its 6thseason, has 12 competing teams, with the Bengaluru Bullscoming out as the winner after defeating the Gujarat FortuneGiants in the final. For the next season, all 12 teams will playeach other twice and the top six teams will make it to theplayoffs, while the bottom four teams will have eliminationrounds. In the 7th season, Bengal Warriors emerged as thechampions after defeating Dabang Delhi. This season also sawmany records being set, such as Pardeep Narwal achieving1,000pointsintheleague,NaveenKumarscoring21consecutive Super 10s, and three raiders reaching 300 raidpoints. Additionally, Neeraj Kumar and Mohit Chillar tied forthemostdefensivetacklepointsinagame.

IVMETHODOLOGIES

A. MultidimensionalScaling

The Multidimensional scaling method was applied to analyzethe performance of kabaddi teams at the league level. Eachround's results were used to calculate the differences betweenteams, and an MDS technique was used to show the teams'performance. To highlight the disparities between the teams,threestrategieswereadopted.Theinitialstrategyrequiredcreatingone MDSandonedissimilaritymatrixforeachround. The second method created a single global MDS chartbycombiningalltheinformationintoadatadissimilaritymatrix. The thirdstrategyproduceda time series for eachteam based on the outcomes of each round. The efficiency oftheteamscanthenbevisualizedusinganMDStechnique using these differences. An example of how to use MDS in thecontextofakabaddileagueseasonisbyconstructingadissimilarity matrix founded on points of every match. Thedissimilaritymatrixcreatedintheanalysisofthenational Kabaddi league reveals the difference between each pair ofteams,withtheentryinthei^{-th}rowandj^{-th}columnrepresenting the difference between team i and team j. ThismatrixisthenprocessedusingMultidimensionalScaling(MDS)toproduceavisualrepresentationoftheteams'performanceintheformofa2Dor3Dplot,wherethepositionofeachteamconveystheir overallperformancethroughoutthegame.Atechniqueforevaluatingthecorrelationbetweenlow-dimensionaldatapoints$y_i$andy_j,which correspond to high-dimensional data points $x_i$and x_j,can be established using a method that involves computingconditionalprobabilities.[7].Todeterminetheconnectionbetween low-dimensional data points y_i and y_j that correspondtohigh-dimensionaldatapoints$x_i$andx_j,adissimilaritymatrix D with dimensions m between the two items is used.This matrix measures the distance between i and j, with d_{ij}beingthevaluerepresentingthatdistance.Theresultisamatrix X, which has a reduced number of dimensions (oftend=1, 2 or 3), that is optimized through the use of gradient descentmethodknownas "teaming". [14].

$$\mathrm{Strain}_D(x_1, x_1, \ldots\ldots, x_N) = \left(\frac{\sum_{i,j} (b_{ij} - (x_i, x_j))^2}{\sum_{i,j} b_{ij}^2} \right)^{1/2} \quad (1)$$

TheMDStechniqueisawaytoconverthigh-dimensional data into a low-dimensional format byreorganizing data points that are similar to be neareach other, while data points that are dissimilar areplaced farther apart. This is achieved by creating amatrix B from the dissimilarity matrix D, and thenusing a double-centering transformation. Output isproduced by calculating the eigen-decomposition ofmatrixB.[10].nThetraditionalMDSalgorithmstarts by computing the square dissimilarity matrixD2fromthedissimilaritymatrixD.Then,itcalculatesthecenteringmatrixJand thematrix.

$$B = -\frac{1}{n} J D^2 J \quad (2)$$

After that, it determines matrix B's d largesteigenvalues and eigenvectors. finally calculates thelow-dimensionalrepresentation matrix,

$$X = E_d \Lambda_d^{1/2} \quad (3)$$

B. K-NNAlgorithm

The The KNN algorithm groups of observations based ontheirsimilaritiestootherobservationsinadataset.Thealgorithm calculatesdissimilaritymeasuresbetweenallobservationstofindthekclosestobservationsofagivenobservation. It then assigns the new observation a categorylabeldependingonthemostcommoncategorylabelofitsk

DOI: rp-9788770040723.178

closestobservations.heKNNapproachalsoappliestomachine learning, where the objective is to locate clusters ofsimilarobservationsorcomparableobservationsthatdifferfrom oneanother.InthecontextofKNN,thealgorithmcategorizesnewo bservationsbasedontheirsimilaritiestoother observations in a dataset. The dissimilarities between allthe observations are calculated, and the k nearest neighbors ofa given observation are determined using these dissimilarities.The class label for the new observation is assigned based onthe most frequent class label among its k nearest neighbors. Inan unsupervised learning scenario, KNN can be utilized todiscoversimilarobservationsorclustersofsimilarobservations based on their differences. This is achieved byperformingclustering,calculatingthecentroid,anddetermining the k nearest neighbors for each observation. Thecentroid is found by taking the average of the cluster. Thelikelihoodofaconnectionbetweentwolow-dimensionalobservations can be determined using a similar method basedon the calculationofconditional probability.

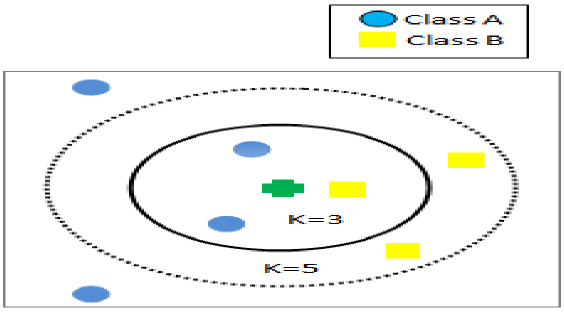

Fig.1.K-NearestNeighnorClassifier

In Figure 1, a diagram depicts the design of a K-NearestNeighbor(KNN)classifier.Thegreenrectangularsphere, representing a test case, will be assigned to differentclassesbasedonthevalueofk.Forinstance,ifkissetto3,the test case will be classified as part of the blue circular class.But if k is increased to 5, the test case will be classified as partof the yellow square class. [8]. A K-Nearest Neighbor (KNN)classifierisusedtoidentifytheclassthatatestsample,represented byx,belongsto.Theclassificationprocessinvolves calculating each group's likelihood based on the K-NN of x, using a specific distance metric, d. The example isassigned to the group with the highest likelihood, which isdetermined by evaluating the distance between x and the knearestneighbors.

C. t-DistributedStochasticNeighborEmbedding(t-SNE).

t-SNEisatypeofnon-linearandself-governingmachinelearningmethodutilizedforsimplifyinghigh-dimensionaldata,suchasimages,audiosignals,ortext,intoalower-dimensional representation. T-SNE aims to maintain the keyinformation contained in the original high-dimensional datawhile condensing it into a more manageable form, usually 2 or3 dimensions. The advantage of using t-SNE is that it cansimplify high-dimensional data into a lower-dimensional form,making it easier to comprehend the relationships between thedatapointsandunderstandthedata'soverallstructure.

Traditional methods of analysing high-dimensional data cansometimes fall short in visualizing it in a way that retains itslocalandglobalaspects,makingt-SNEavaluabletool inthese situations. [3]. The t-SNE algorithm is used to deal withthe problem of visualizing high-dimensional data effectivelywhilepreservingbothlocalandglobalstructures.Ittransforms the distances between high-dimensional data pointsintoprobabilitiesofsimilaritythroughaprocesscalledStochasticNeighborEmbedding.Thisresultsinalower-dimensional representation that retains important informationfrom the original data. The t-SNE technique is a widely usedtoolinfieldssuchascomputervision,naturallanguageprocessing,andbioinformatics,asithelpsconverthigh-dimensional data into a more manageable and interpretableform. This allows for deeper insights into the structure andrelationships within the data, making it a valuable tool for dataanalysisandexploration.[5].

$$p_{j|i} = \frac{\exp\left(-\|x_i - x_j\|^2/2\sigma_i^2\right)}{\sum_{k\neq i}\exp\left(-\|x_i - x_k\|^2/2\sigma_i^2\right)} \quad (4)$$

The t-SNE algorithm aims to compute a similar conditionalprobability between the low-dimensional data points y_i and y_j,whichcorrespondtothehigh-dimensionaldatapoints$x_i$andx_j.Thedifferencebetween$x_i$and$x_j$isrepresentedasthemagnitude of their separation $||x_i-x_j||$, and the spread of datapoints in high-dimensional space is represented by σi. Theobjective is to compute the straight-line distance between $y_i$andy_j,whichisrepresentedasthemagnitudeoftheirseparation$||y_i-y_j||$,inordertocalculatetheconditionalprobability between the low-dimensional data points. The aimis to find a way to calculate the relationship between the high-dimensional feature vectors of each Kabaddi teamandthecorresponding low-dimensional data points, while preservingthesimilaritiesbetweentheteams.Todothis,t-SNEcalculatestheprobabilitiesbasedonthedistancesandvariancesbetweenthehigh-dimensionaldatapointsandminimizes the difference between these probabilities and thecorrespondinglow-dimensionalpoints.Thisprocessallowsforareductionofthehigh-dimensionaldataintoamoremanageablelower-dimensionalspacewhileretainingimportant information about the teams' performance, such astheir record of wins and losses, the points they score, and thestrategies they use for defense. By using t-SNE, it is possibletogainvaluableinsightsintotheperformanceofKabadditeams,whichcanaidcoaches,players,andanalystsinunderstanding thestrengthsand weaknessesofeach team.

IV. RESULTANDDISCUSSION

Theoutcomesofthestudyindicatedthatutilizingt-DistributedStochasticNeighborEmbeddingandMultidimensionalScalingasameansofreducingthedimensions of the data was effective in comprehending theintricaciesofthegameofKabaddiandidentifyingsignificant

International Conference on Recent Trends in Data Science and its Applications
DOI: rp-9788770040723.178

variables that influence a team's success. The t-DistributedStochasticNeighborEmbeddingmethodwasparticularlyeffectiveinvisuallyrepresentingcomplexandnon-linearrelationshipsinthedatabycreatingareduced,simplifieddepiction of the team's success over time. In Figure 2. Showthe team who won the match according to the win points therearefourclusters,cluster1representsthehigherwinningpoints above13to17,andcluster2denotesthepointsbetween 12 to 14. After which, the winning point from 6 to 10comesunderclustering3.Finally,cluster4representstheremainingpoints.Infigure3.Showsthetwo-teamperformance analysis within the season 2017-2018. The teamU Mumba Played the game in two seasons and score the highpointsitisrepresentedinfigure 3.

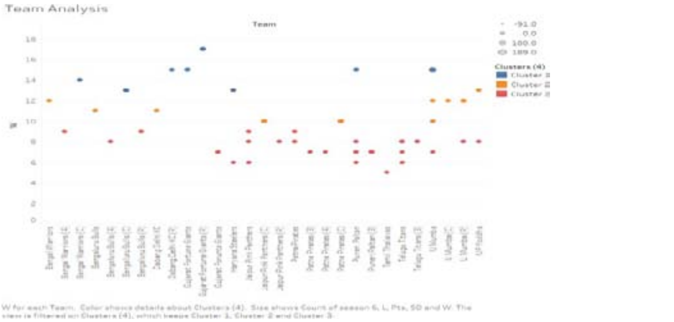

Fig.2.TeamClustering

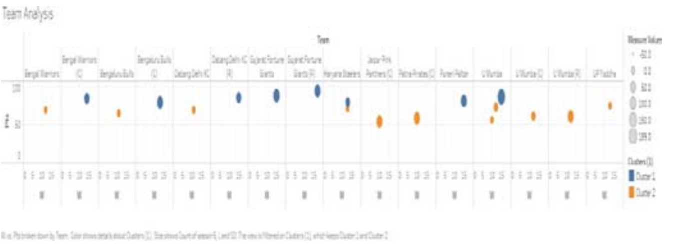

Fig.3.IllustratetheTwoteamperformance

The resulting scatter plots, as seen in Figs. 4 and 5, clearlyexplain the teams' season-based dissimilarity matrix. The K-NNalgorithmfaredwellregardingclustering,withthemaximumaccuracyattainedwhileutilizingt-SNEasthedimensionalityreductionmethod.Thestudy’sresultsdemonstratedthatusingt-DistributedStochasticNeighborEmbeddingandMultidimensionalScalingasameansofreducingthedimensionsofthedatawashelpfulincomprehending the intricacies of the game of Kabaddi andidentifying crucial factors that impact a team's performance.Thet-DistributedStochasticNeighborEmbeddingmethodsuccessfullycreatedareduced,simplifiedrepresentationofthe team's performance over time. The scatter plots produced,shown in Figures 4 and 5, effectively displayed the team'sseason-based differences. The K-Nearest Neighbors (K-NN)algorithmperformedwellingroupingthedata,withthehighestaccuracyachievedwhent-SNEwasusedasthemethodofsimplifyingthedimensionsofthedata.This highlightstheeffectivenessoft-SNEinpreservingtheessentialstructureoftheinformationwhilereducingitscomplexity. The team performance during the seasons from2017 to 2018 is shown in Figures 4 and Figure 5 displays theteam's performance using Multidimensional Scaling (MDS)and the data points are relatively close to each other. Figure 3shows the results using the t-Distributed Stochastic NeighborEmbedding (t-SNE)technique.

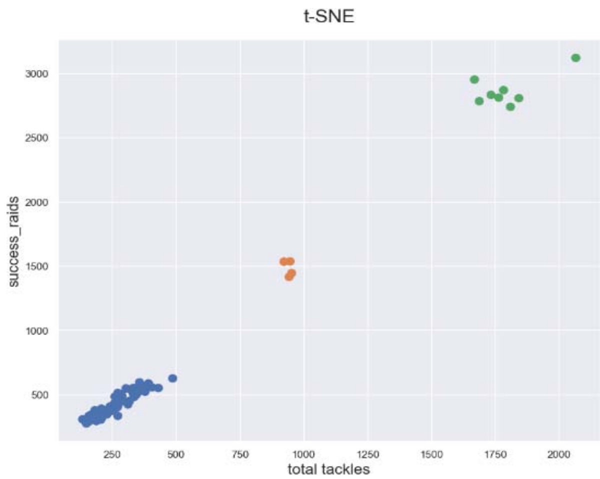

Fig.4.t-DistributedStochasticNeighborEmbedding(t-SNE)

Both algorithms display the team's performance based on theseasons.MDSandt-SNEwereusedtotransformthedissimilaritydatapointsintosimilaritymatrices.t-SNEeffectivelyanalyzesthecomplexconnectionsbetweenthedatapoints,andcomparedtoMDS,itspeedsupthecalculationwhilestillproducingaccurateresults.Therepresentation of the team performance over time can be seenin t-SNE, which also visualizes the intricacies of the game ofKabaddi.InFigure4.Showthepairwisedistancesbetweenthe data points in the high-dimensional space in the lower-dimensional space and preserve it. In kabaddi data, its takeswon points and calculate the Euclidean output space where thedistances between the points are proportional to the distancesbetweenthepointsintheoriginalhigh-dimensionalspace.Figure5shows topreservethelocalstructureofthedatapoints by minimizing the divergence between two probabilitydistributions. By using t-SNE the won data points preserve thestructureofwondatapointsstructure,andtheresultisadiscrete,embeddedoutputspacethatcapturesthelocalrelationshipsbetween thewondatapointbetweentheteams.In conclusion, our analysis showed that using t-DistributedStochasticNeighborEmbedding,MultidimensionalScaling,andK-NearestNeighbors(K-NN)togetherwasausefultechniqueforunderstandingtheintricaciesofthegameofKabaddi and identifying important factors that have an impactonteamsuccess.Furtherresearch,usingalargerdataset,could bedoneto validate thesefindingson alarger scale.

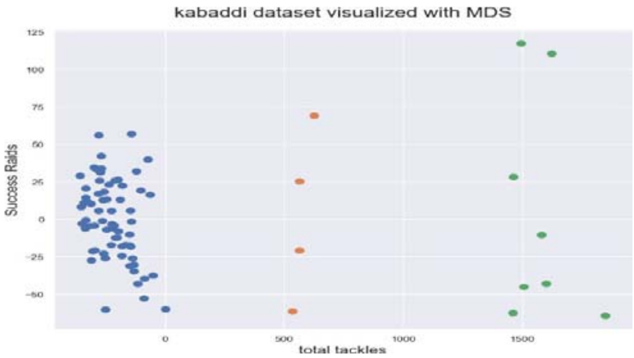

Fig.5.Multi-DimensionalScaling

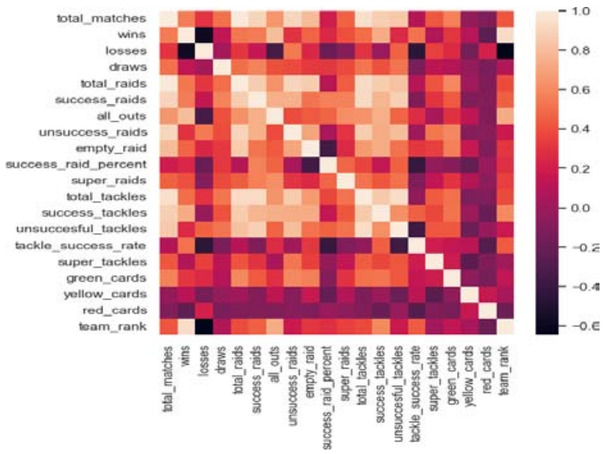

Fig. 6.Multi illustrates the correlation and co-variance of the dissimilaritymatrixintheKabaddidatasetthroughaheatmap.

Fig. 7.Representation of Tree map- it shows which team gets high and lowwon points in a season form 2017-2018, the dark colour shows the high pointsandlightcoloursshows theleast scoringpoints

V. CONCLUSIONS

In order to understand Kabaddi effectively, this researchcontributes in analysis of a team's performance over severalseasons.MDSandt-SNE,twodatavisualisationmethods,were applied to accomplish this. These techniques made iteasier to evaluate the progress of the team and determine thekeyfactorsthataffectitssuccess.Theanalysis'sfindingsprovided a crucial new understanding of the dynamics of thegameandthevariablesaffectingateam'sperformance.The study was able to offer a thorough insight of Kabaddi and itscomplexities by evaluating the MDS and t-SNE data. Teamsand coaches who desire to perform better might take use ofthese insights. They can design strategies and techniques thatwill enable them to excel in the game by being aware of theimportant success factors. Overall, the study has added muchto our understanding of Kabaddi and is likely to be of greatinterest to fans of the game. Future research could considerusingmachinelearningtechniquessuchasclustering,classification, and regression to identify patterns and makepredictions based on the data applying the findings to practicalapplications such as team selection, training strategies, andgametacticstoimprove performance.

.

REFERENCES

[1] A. Chatzimparmpas, R. M. Martins and A. Kerren, "t-viSNE: InteractiveAssessmentandInterpretationoft-SNEProjections,"in*IEEETransactions on Visualization and Computer Graphics*, vol. 26, no. 8,pp.2696-2714,1Aug.2020,doi:10.1109/TVCG.2020.2986996.

[2] VanDer Maaten L.Accelerating t-SNEusingtree-basedalgorithms.TheJournalofMachineLearningResearch.2014Jan1;15(1):3221-45.

[3] Van der Maaten, Laurens, and Geoffrey Hinton. "Visualizing data usingt-SNE."*Journalof machinelearningresearch*9,no.11 (2008).

[4] Sitharthan, R., Vimal, S., Verma, A., Karthikeyan, M., Dhanabalan, S. S., Prabaharan, N., ...&Eswaran, T. (2023). Smart microgrid with the internet of things for adequate energy management and analysis.Computers and Electrical Engineering, 106, 108556.

[5] Devassy, Binu Melit, and Sony George. "Dimensionality reduction andvisualisation of hyperspectral ink data using t-SNE." *Forensic scienceinternational*311(2020):110194.

[6] Anowar, Farzana, Samira Sadaoui, and Bassant Selim. "Conceptual andempirical comparison of dimensionality reduction algorithms (pca, kpca,lda, mds, svd, lle, isomap, le, ica, t-sne)." *Computer Science Review* 40(2021):100378.

[7] Muller S.T.Distance, similarity, and multidimensional scaling (2019)https://pages.mtu.edu/ shanem/psy5220/daily/Day16/MDS.htmlGoogleScholar

[8] Sakib,Shadman,MdAbuBakrSiddique,andMdAbdurRahman."Performance evaluation of t-SNE and MDS dimensionality reductiontechniques with KNN, ENN and SVM classifiers." In *2020 IEEE Region10Symposium(TENSYMP)*,pp.5-8.IEEE,2020.

[9] Borg,Ingwer,andPatrickJFGroenen.*Modernmultidimensionalscaling: Theory and applications*. Springer Science & Business Media,2005.

[10] Wickelmaier,Florian."AnintroductiontoMDS."*SoundQualityResearchUnit,AalborgUniversity,Denmark* 46,no.5(2003):1-26.

[11] Buja, Andreas, Deborah F. Swayne, Michael Littman, Nathaniel Dean,andHeikeHofmann."XGvis:Interactivedatavisualizationwithmultidimensionalscaling."JournalofComputationalandGraphicalStatistics(2001):1061-8600.

[12] Tzeng,Jengnan,HenryHorng-ShingLu,andWen-HsiungLi."Multidimensionalscalingforlargegenomicdatasets."BMCbioinformatics9(2008):1-17.

[13] Moshika, A., Thirumaran, M., Natarajan, B., Andal, K., Sambasivam, G., &Manoharan, R. (2021).Vulnerability assessment in heterogeneous web environment using probabilistic arithmetic automata. IEEE Access, 9, 74659-74673.

[14] Njock, Pierre Guy Atangana, Shui-Long Shen, Annan Zhou, and Hai-MinLyu."EvaluationofsoilliquefactionusingAItechnologyincorporatingacoupledENN/t-SNEmodel."SoilDynamicsandEarthquakeEngineering130(2020):105988.

DOI: rp-9788770040723.178

[15] Pezzotti, Nicola, Boudewijn PF Lelieveldt, Laurens Van Der Maaten,ThomasHöllt,ElmarEisemann,andAnnaVilanova."Approximated andusersteerabletSNEforprogressivevisualanalytics."IEEEtransactions on visualization and computer graphics 23, no. 7 (2016):1739-1752.