The application of the Gaussian Mixture Model algorithm for the unsupervised analysis of surface topographies

Marcin Hinz[a], Doha Meslem[b], and Stefan Bracke[c]
[a] University of Wuppertal, Wuppertal, Germany, m.hinz@uni-wuppertal.de
[b] University of Wuppertal, Wuppertal, Germany, dmeslem.dm@gmail.com
[c] University of Wuppertal, Wuppertal, Germany, bracke@uni-wuppertal.de

Abstract: Unsupervised learning is a type of machine learning that deals with analyzing provided data to draw its patterns without pre-expected results. The patterns are drawn by clustering the data. In this paper, the Gaussian Mixture Model (GMM) is chosen as the clustering method, and a parameter-study is performed. This parameter study includes pre-processing the data using standardization and normalization techniques, along with tuning GMM parameters. This method undergoes a parameter analysis, to find the set of parameters that gives out the clusters that are most compatible with the manufacturer's set criteria, in this case, roughness values Ra. Moreover, the data is clustered in an unsupervised approach and upon studying a wide range of parameter possibilities, the algorithm is expected to provide an output fitting a certain prescribed standard (Ra), mimicking therefore a supervised approach without having to train the data, but by rather studying the patterns. The paper briefly discusses the theory behind the methods used based on published research, but the main scope of the paper is the feasibility of the method.

1. INTRODUCTION

The cutlery industry is well established yet continuously adapting to modern requirements. The products are being continuously optimized, improved and refined. Such changes are however context-based. Moreover, the products are manufactured with the focus of them being individualized, customized solutions based on the expected usage. Consequently, changing any of the manufacturing process parameters or the materials changes the surface topography, which is the most important quality for these products.

Manufacturing processes of fine grinded surfaces – cutlery, are complex and consist of more than 50 steps [1]. They are controlled by many production parameters, such as the feed rate, cutting speed, grinding disc, cutting fluid, contact force and process time. The surface topography is then measured by the following target parameters: gloss, roughness, and colour. To achieve the sought-for surface, a compromise needs to be reached. This compromise depends on which characteristic of the product is more important for the specific task or product.

For analysing the quality of the product, and hence of the accompanying optimization technique, unsupervised learning methods are used in this study. The application of the Gaussian Mixture Model (GMM) algorithm for the unsupervised analysis of surface topographies is proposed as a solution.

The solution discussed in this paper covers the following aspects: firstly, an analysis of a real data set out of the cutting industry and comparison of the achieved results to the company targets. Secondly, the company results are set using the specification limits (upper and lower) and for that, the ideal products are defined from the manufacturer point of view. Finally, a comprehensive analysis of the parameters of the algorithm is performed. This is achieved by studying and setting the proper set of parameters within the algorithm.

This paper is divided into two main parts, the theoretical and the practical parts. The theoretical part acts as transition between the research question and the practical part, where the question is answered. It consists of general background information of machine learning and a thorough explanation of the GMM algorithm. The practical framework consists of the parametric analysis of the algorithm, an explanation of the data, a display, and a discussion of the obtained results.

2. Literature Review

GMM is an area of active research. This is well founded by the amount of the literature present in its regard. As previously mentioned, this paper uses gaussian mixture model as the unsupervised learning method to analyse the surface topographies. Numerous literature discuss unsupervised machine learning clustering using GMM, the theory behind it and how to program it. Andriyanov, Tashlinsky, and Dementiev in their article [2] discuss gaussian models clustering in detail. A further example is McLachlan and Rathnayake's article discussing the number of components in GMM [3].

Due to its "effectiveness and efficiency", GMM is researched a lot [4]. Many relevant literature however focuses on the mathematical aspect of GMM. While it is justified that the focus of researchers be concentrated on this aspect due to its complexity and huge potential, it still serves as the preliminary steps that saturate the theory.

Moreover, literature here divides into more than one aspect, mainly explaining the *mathematics* behind GMM, Expectation-Maximization (EM) iterations, and developing the current GMM method. The developments include for example aspects where the current algorithm is not fully explored, such as using it with incomplete data, as discussed by Zhang et al. in their article [4]. Another aspect is the dimensional representation of data sets, as explored by Shrivastava and Tsui in their article [5]. Liu, Cai and He also propose an improved gaussian mixture model documented in their article [6]. They introduce a GMM approach with local consistency. Those research papers are some of many examples of papers investigating areas where GMM could be adapted. All of this occurs by applying the algorithms and the improved versions on already-present data, to study them and the mathematics behind them rather than the data itself, hence focusing on theory and understanding rather than application.

Several literature take this however, a necessary step further. A focus is put instead on the study of experimentally acquired data, such as that done by Selwyn and Difranco in their study [7]. Yussof et al., and Ye et al.'s papers to name a few, are also article-archetypes of clustering self-acquired data using gaussian mixture models in various applications; biological, transportation and traffic, as well as telecommunications respectively [8, 9].

All such articles draw the same conclusion, that GMM is a well-fit clustering-method for analyzing their data. This explores the applicability of GMM, it however stays within the premises of explaining and analyzing a situation rather than solving a problem.

Nevertheless, there exist multiple pieces of literature that serve that purpose. Within this scope comes condition monitoring using GMM. Zorriassatine et al. in their case study [10] and Chauhan and Surgenor's [11] cover condition monitoring by fault detection.

Additionally, condition monitoring is also covered in many fields. Yu covers it in the field of Machine Tools [12], Qui et al. in the field of Aircraft by studying the damage propagation [13], and Shen et al. in the field of Automotive Engines [14] to name a few. Yusoff et al. in their article [15] and Heyns, Heyns and De Villiers [16] also discuss condition monitoring using GMM in their respective fields, namely: turbomachinery and the monitoring of a gearbox.

However scarcely in any literature is it researched on surface topographies using GMM. This paper serves as to abridge the gap present in the practical aspect by enriching the current literature concerning feasibility of unsupervised machine learning in condition monitoring. Moreover, this paper does not add to the present literature in quantity by multiplying the scarce practical literature, but by a introducing a new perspective with which unsupervised learning can be approached as will be shown later in the paper. Additionally, the methods used in this paper are also unprecedented within the contexts they are used in. The generic character of this research allows it to be applied to other sets of extracted data of fine grinded surfaces.

3. Theoretical Background

In this chapter, the theoretical background needed for the understanding of the application of the algorithms as well as their parametrization is briefly described.

3.1. General theory of machine learning

Machine learning consists of three main types along with their derivatives or combinations. Mainly, supervised, unsupervised, and reinforcement learning. Derivatives include semi-supervised, self-supervised, and multi-instance learning techniques.

The main difference between the main machine learning types lies within which data they can be used on and what is the expected outcome of the process. Supervised machine learning is applied on labelled data where the result is predetermined, and the algorithms work on the data to give an expected output. Algorithms therefore are continuously trained, tested, and corrected until an acceptable level of accuracy is reached.

Unsupervised learning is applied on unlabelled data, where the output is unknown, and the algorithm is expected to draw conclusions from the data based on patterns between the datapoints and result in an output. This is therefore reliant on how the algorithm is built/programmed and on how it processes the data points. This however is not random. Supervised and unsupervised machine learning types distinguish and recognize patterns between datapoints in specific ways. Supervised machine learning methods determine the output by classification and regression. Unsupervised machine learning types can on the other hand determine outputs by clustering or dimension reduction. Clustering is nonetheless the more common method.

The following methods are the most prevalent in clustering: exclusive-distance based, overlapping, hierarchical-based on similarities or differences, density-based, association and probabilistic. Each of these methods can be applied using various algorithms.

Each algorithm has a specific parameter-set, method and code based on the expected result. The algorithms are also under continuous research, testing, and improvements [17, 18].

3.2. Gaussian Mixture Model

The gaussian mixture model is an unsupervised machine learning model based on a probabilistic approach. It clusters the data by associating each data point with a gaussian distribution curve. Moreover, this model finds out the probability of each data point to correspond to each of the clusters, and the gaussian with the highest probability is the cluster that the datapoint belongs to.
The model samples the data as a mixture of gaussians. This means that the data is thought to be fully explained by a set of multiple gaussian graphs, a mixture that is. The gaussians here are latent variables, they are unknown and are to be found.

Modelling large sets of data using GMM is plausible because of the central limit theorem. This theorem states that the larger the number of data, population, or points, the more it mimics a gaussian distribution.

The gaussian probability distribution is described using two main parameters, the mean and the variance. The mean describes the centre of the distribution (focus), where most of the data are located, for a standard non-skewed gaussian while the variance describes the spread of the data away from the mean (scatter).

To cluster multivariate data, a multivariate gaussian distribution is needed. A multivariate gaussian is however specified by a vector of mean and a vector of variance corresponding to the number of variables. A multivariate variance is called covariance, and it is a matrix.

As stated, the gaussians are described by their parameters, and the data in the gaussian mixture model are specified by the latent gaussians. This means that obtaining which gaussians cluster which data points can be done by either finding the parameters mean and covariance, or by finding the related latent gaussians. Thus, either one or the other needs to be given. However, for this unsupervised machine learning algorithm, the only present information is the data. For this purpose, the gaussian mixture model uses an expectation-maximization method.

EM is an iterative method used to find the parameters and consequently the corresponding gaussians. Additionally, it finds each of the probabilities of each of the datapoints to each of the gaussians and assigns them accordingly. This iteration method always converges, which is a major advantage of the algorithm.

The iteration method EM consists of two steps: expectation and maximization. Both steps make up one iteration. Initially in the E-step, the algorithm allocates random mean and covariance values for each of the multivariate gaussians, to find a multivariate gaussian density. The mean, covariance, and weights are fixed while the probability of each datapoint belonging to each data point is found. The number of gaussians is predetermined as the number of components. This is done using the following equation [19]:

$$p_k(x \mid \theta_k) = \frac{1}{(2\pi)^{d/2}|\Sigma_k|^{1/2}} e^{-\frac{1}{2}(x-\mu_k)^t \Sigma_k^{-1}(x-\mu_k)} \quad (1)$$

$$\theta_k = \{\mu_k, \Sigma_k\} \quad (2)$$

Where $\mu_k \; and \; \Sigma_k$ are the mean and covariance respectively, k is the number of components and x is the data point.

In the M-step, the weights, mean, and covariance parameters are updated using the following equations while the probability is fixed. The updated values take account of the weighted contributions of the datapoints from the previous step [20].

$$\lambda_k^{new} = \frac{N_k}{N} \quad (3)$$

$$\mu_k^{new} = \frac{1}{N_k} \sum_{n=1}^{N} w_{nk} \cdot x_n \quad (4)$$

$$\Sigma_k^{new} = \frac{1}{N_k} \sum_{n=1}^{N} w_{nk} \cdot (x_n - \mu_k)(x_n - \mu_k)^T \quad (5)$$

Where w_{nk} is weight of the data point, N is the total number of datapoints, N_k is the sum of the weights for the k[th] cluster and λ_k^{new} is the updated mixture weight.

The GMM uses the EM method to fit and predict the labels from the data. It fits the data as many times as stated by the number of initializations. Number of maximum iterations determines the number of times the EM steps iterate within one run of number of initializations. The model keeps iterating till either the number of maximum iterations or a stopping criterion is reached. If the change of the likelihood function is less than the tolerance, the iterations stop. [20]

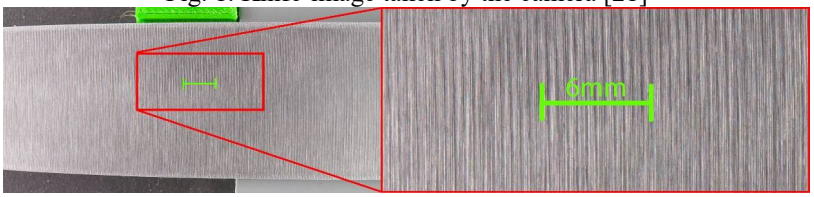
4. Implementation and parametrization of the algorithm

The practical aspect of the paper, where the GMM is applied on the set of data described in the following section, consists of two main objectives. Mainly, a comprehensive and structured testing of various parameter setups to find the best parameters and a quantitative comparison of the achieved results.

4.1. Data sources / experimental rig including feature extraction

Three types of knives were provided by the manufacturer with different surface-topographies. The data used is therefore taken from a variety of cutlery samples and divides into two main parts, mechanical and optical. The former consists of color, roughness, and gloss values. Optical values on the other hand consist of a variety of line and optical measurements analyzed with the help of computer vision. Only a part of the surface is measured and clearly specified. As a variable control, all pictures are cropped in the same measurements and the reference part of the cutlery is consequently identified. [21]

Fig. 1. Knife-image taken by the camera [21]



The optical values are data collected about the knives to define their current state. The mechanical values are on the other hand considered as target values, with which the final surface topography and hence the characteristics are defined. The optical data is therefore intended to be eventually compared to the mechanical data to check for possible compatibility.

The collected data on which the methods are applied is multivariate, with 42 variables (features), structured, numerical and is taken from the surfaces of the three knife-types. Of the three target-variable-categories, the one of most interest in this study is roughness represented by the Ra value. These are calculated for each knife using the following equation [22]:

$$R_{\mathrm{a}} = (1/L_{\mathrm{m}}) \int_{x=0}^{x=L_{\mathrm{m}}} |y| \mathrm{d}x \qquad (6)$$

Where L_{m} is the total scanned length in the x direction.

4.2. Numerical setup in detail

For the purpose of the study, the optical data is first pre-processed. This is done using two standardization techniques, one normalization technique and a combination of those. The order of combining is also varied, yielding seven pre-processed mechanical data combinations and one is left without pre-processing. Standardization is rescaling the distribution of the values to change the mean to 0 and the standard deviation to 1. It is done by subtracting the mean and dividing by the standard deviation. While the two normalization techniques are the MinMax and the normalizer functions from the Sklearn library. They rescale the data to a range between 1 and 0. [23, 24]

For the covariance parameter, there are four types that this algorithm uses. These types are “full”, “tied”, “diag” and “spherical”. They dictate the shape of the clusters and are specified by their matrices, consisting of standard deviations and their correlations. For the covariance type “spherical”, the standard deviations on the diagonal are equal while the rest is zeros, showing no correlation between the variables. This spread has one variance. This gives therefore spherical clusters of the same shape but different sizes. In the covariance matrix of type “diag” the standard deviations on the diagonal vary while the rest of the matrix is also zero. This gives diagonal gaussians of different sizes. Type “Full”

can on the other hand take any of the shapes and has therefore varying standard deviations and non-zero values on the rest of the matrix, giving varying relationships. The gaussians here are of different shapes and sizes. The type "tied" ensures all gaussians take the same shape and standard deviation, giving therefore equal clusters in shape and size. [25]

The pre-processed data can then be used by the GMM algorithm. The parameter lists are initially set. The following parameters are tuned: the covariance types, number of initializations, number of iterations and tolerances. Number of clusters or components is set to three. For the covariance types, all the possible types are chosen. For number of initializations and iterations four values are set, namely: 100, 200, 500 and a 1000. These are chosen to cover a wide range to check the trends they create and study which tendencies they tend to follow. For tolerances two values are chosen, 0.01, 0.0001. Based on multiple runs of the algorithm on several different pieces of the data using a variety of tolerances ranging from 0.01 till 0.00001 it was determined that those are sufficient. Moreover, the efficiency values with tolerances of 0.0001 and 0.00001 yielded a discrepancy of around 1% and same happened with tolerances of values 0.01 and 0.001. for this reason, only one of each of the two sets were chosen, and the higher the tolerance value the faster the convergence occurs, which saves the computational effort and at the end the time needed for the calculations.

All the parameter-combinations are run by the algorithm. Each set of parameters is run through the algorithm and then fitted to result in an array of 0, 1 or 2 corresponding to which cluster each data point belongs. The target variable of interest in this paper as previously mentioned is the roughness value classified in three classes. This is chosen because of the manufacturer set criteria. The manufacturer provided a set of roughness values forming an upper and a lower limit. Values of roughness not within this range are considered from the manufacturer's point of view of no use and are therefore avoided in the production of the knives. The clustered data are consequently set to be compared to this set with the upper and lower limit. Because of this comparison, the number of clusters is chosen to be equal to the number of classes corresponding to the roughness values set by the manufacturer.

The clusters' order needs to therefore comply with roughness order. As the GMM algorithm labels clusters arbitrarily, yet consistently nonetheless, the order of the clusters may need to be redone. This is done by renaming the clusters. Renaming the clusters occurs on basis of the roughness values present in each of the clusters. Moreover, the mean of the roughness values corresponding to each cluster is calculated. The clusters are then renamed in ascending order based on the roughness mean. The lower the mean, the lower the cluster number, that is.

The relabelled clusters are now ready for comparison. The comparison shows how well the clustering method clusters the data in correspondence to the manufacturer's set criteria. The efficiency of the algorithm is measured by various counters, the most important of which is the efficiency. The best set of parameters with the highest efficiency is the set that is supposed to be chosen to analyse the quality of the surface topographies of knives within this criterion and this production process.

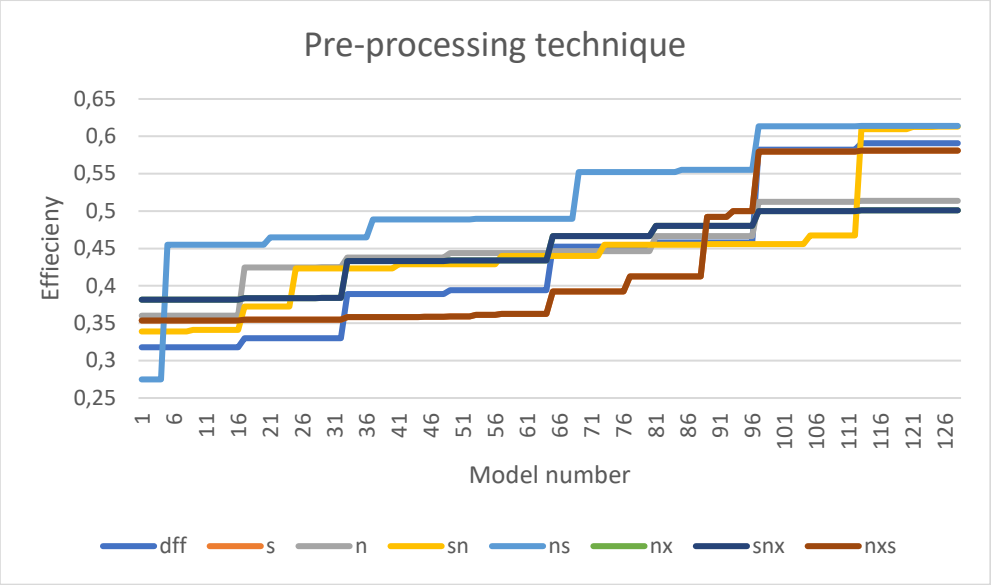
The generic character of this experiment allows it to be used on other manufacturer set criteria, with different production processes and in different settings. As previously shown, the data is unlabelled, and is clustered using unsupervised learning methods. By comparing it to some prescribed set of data and expecting it to comply with it, by calculating the efficiency, and 'training' it, by setting a GMM-parameter analysis, this study mimics a supervised machine learning method. Therefore, in this study, an unsupervised machine learning method simulates, in some aspects, supervised machine learning, which adds to the importance of such literature in the field.

5. Discussion of results

This research covers a comprehensive analysis of the parameters of the GMM. The parameter-combinations are run through the algorithm and the efficiency attained by each of the sets is recorded. The parameters are then analysed separately as seen by the graphs below and plotted against their

efficiencies. This allows for a comparison between the values chosen. The best set of parameters is recorded, and an analysis is done.

Fig. 2. Pre-processing techniques and efficiency.



Data with no pre-processing is initially run once. Subsequently, seven techniques were used; standardization, normalization using normalizer, standardization followed by normalization, normalization followed by standardization, normalization using MinMax normalizer, standardization followed by MinMax normalizer, and MinMax normalizer followed by standardization respectively.

The normalized-standardized technique gives the lowest efficiencies at certain parameter-combinations, they however proceed to give the highest efficiencies in all other combinations, in comparison to all other pre-processing techniques. It also covers all the efficiency range along various parameter-combinations. Data pre-processed using standardization-MinMax normalization has its lowest efficiency higher than that of all other techniques. Its highest efficiency is however only around 10% more, giving overall the smallest range. 70% of the Standardized-normalized data gives an efficiency of around 45%, while the other 30% is unequally distributed with 20% giving higher efficiencies reaching up to the overall maximum. Around half of the data pre-processed using MinMax-standardization give a stable 35% efficiency, which is the lowest value attained by this technique. Additionally, around 50% of the data give the overall lowest efficiencies compared to all other techniques, ranging between 35 and 40% efficiency. Normalized data and standardized-MinMax data behave similarly with an increase and decrease at the same parameter-combinations.

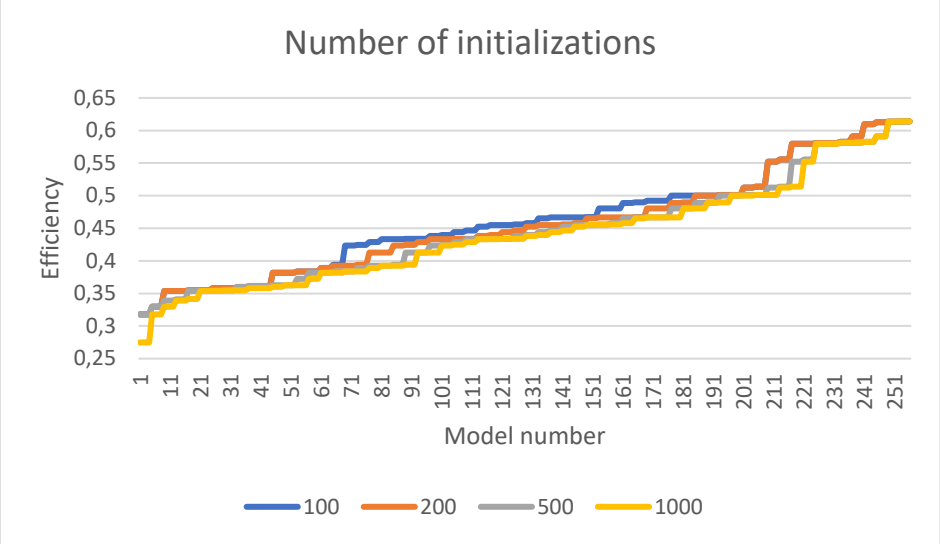
In general, however, none of the techniques always results in the lowest efficiencies. There always exist combinations that increase the efficiencies, unlike the trend shown by normalization-standardization where it takes over the overall highest efficiency. It is also noteworthy to state that the order of the techniques largely influences the output.

Standardized-MinMax and MinMax behave as replicas in all parameter-combinations, and so do MinMax- standardized and standardized data.

Pre-processing the data is also not always necessary. While the highest efficiencies are given by double pre-processing, unprocessed data still gives high efficiencies of around three percent less than the maximum. This is relevant if the computation time is a decisive factor. This results with the combination of “full” covariance type, 0.0001 tolerance value, and all combinations of the number of iterations and initializations.

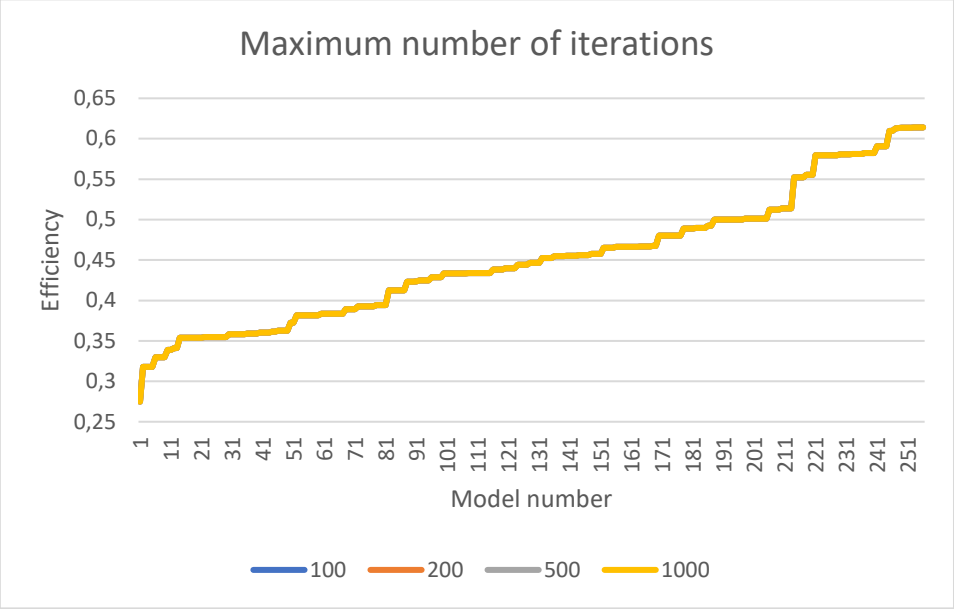
If the pre-processing technique is predetermined, these combinations should be used: for normalization, standardization, MinMax-standardization, tolerance of 0.01, covariance type “full” for all initializations and iterations. For normalization-standardization and without pre-processing, same combination except for tolerance, which is 0.0001. for standardization-MinMax and MinMax the same initial parametric combination but “diag” instead of “full”. Standardized-normalized also uses the same combination but here the number of initializations becomes relevant, with 100 and 200 giving the highest efficiency.

Fig. 3. Number of initializations and efficiency.



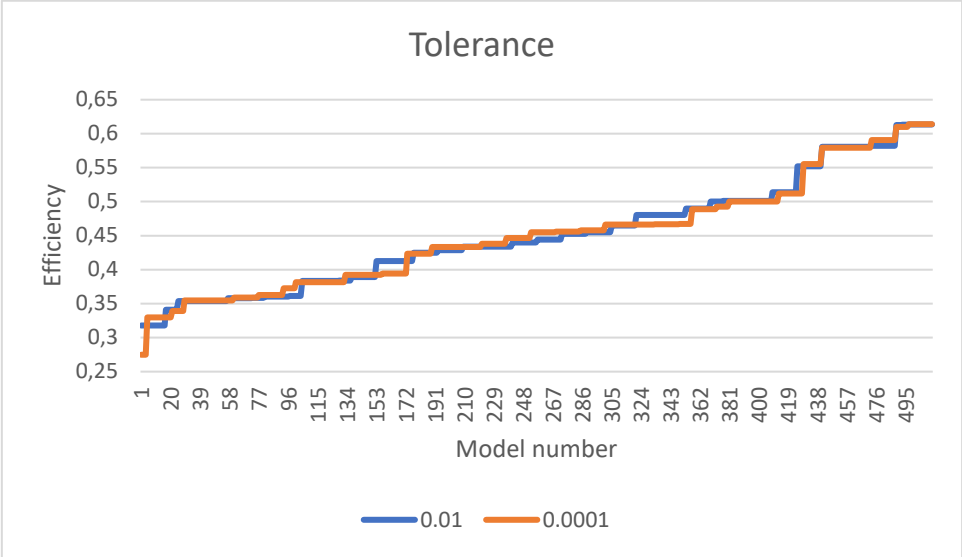
The number of initializations as stated earlier also does not have a large effect on the efficiency. The lowest and the highest efficiencies are achieved by 1000 initializations meaning that the combination has the determining effect. Nonetheless, around 40% of the model combinations give slightly higher efficiencies with 100 initializations in comparison to the three other values.

Fig. 4. Maximum number of iterations and efficiency.



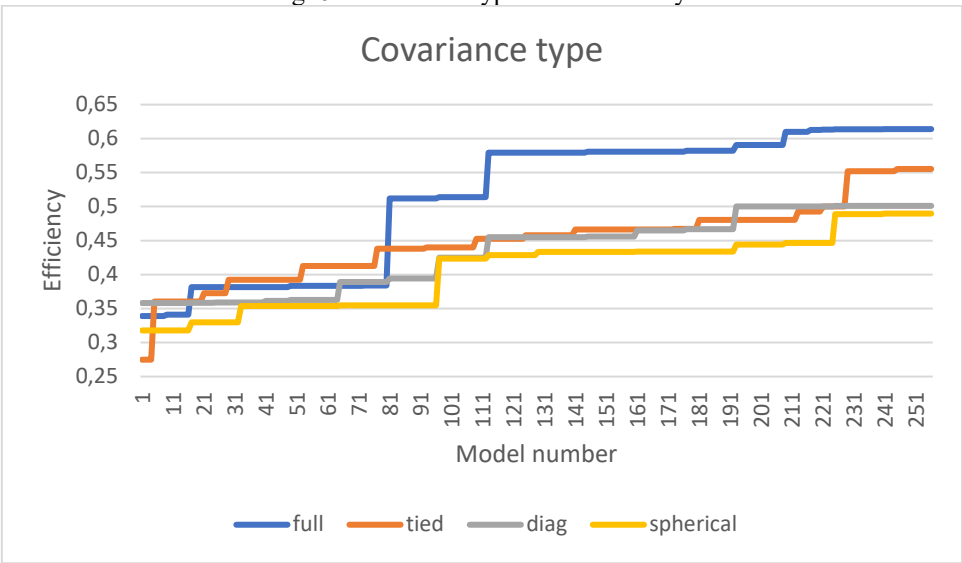
Maximum number of iterations does not affect any parametric combinations. Meaning that convergence occurs, and the number of iterations is not the stopping criterion here.

Fig. 5. Tolerance and efficiency.



The overall lowest efficiencies are achieved by the lower tolerance 0.0001, and the lowest efficiency the tolerance value 0.01 achieved is around five percent higher than that.

Fig. 6. Covariance types and efficiency.



The “full” covariance types give significantly higher efficiencies overall, with the maximum being given by “full”, followed by “tied” with five percent less. The lowest efficiency is achieved also by “tied”. While the “spherical” covariance type consistently gives the lowest efficiency in comparison to all other types. The lowest efficiency attained by the “diag” type is 36% which is higher than that of all other types.

Several parameter sets achieved the highest overall efficiency. The highest efficiency was achieved for “full” covariance type, tolerance of 0.0001 and a pre-processing method of first normalizing the data and then standardizing them. Number of iterations and initializations however varied. All combinations of number of iterations and initializations, given all other parameters fixed as mentioned above, gave the same value of efficiency. This shows that these two parameters did not affect how well the algorithm works on the data and are in comparison insignificant.

With around 50% of the highest efficiency, the lowest efficiency was attained by the following parameter sets: normalization followed by standardization as the pre-processing technique, a tolerance

of 0.0001 and all values of the maximum number of iterations. The covariance type is however 'tied', and the number of initializations is 1000. The first three parameters in these sets were common with the sets for the highest numbers of efficiencies. In this context, this means that the covariance type and the number of initializations are the more determining factors.

For the sake of completeness, the parameters that most affect the efficiency of the algorithm to analyze the surface topography are mainly the pre-processing techniques, followed by the covariance types, then by the tolerance.

6. CONCLUSION AND OUTLOOK

This paper studies the usage of gaussian mixture model unsupervised machine learning algorithm on data extracted from the surfaces of three knife-types provided by the manufacturer. The data extracted using computer vision is clustered using this probabilistic approach and then compared to certain roughness values also provided by the manufacturer. The manufacturer-set criteria act as decisive factors on the usability of the knives. The Model studies the surface topographies and analyses their quality. This is done using a parameter-study, including five of the gaussian mixture model parameters, mainly, the maximum number of iterations, the number of initializations, pre-processing techniques, tolerances, and the covariance types. These are varied and the clustered outcome of each set is compared to the manufacturer-set upper and lower limits. The efficiency of each set is then calculated, and the results are compared and discussed.
In the upcoming research, a set of different unsupervised algorithms will be analysed and discussed in detail. Based on several algorithms, the overall efficiency as well as the best algorithm for the purpose of similar analysis with similar data shall be estimated.

References

[1] M. Radetzky and S. Bracke. "*Manufacturing process parameter adaptation for high precision fine grinded surface topographies: Customised approach within cutlery case study*", Procedia CIRP, 96, pp. 319-323, (2021).

[2] N. Andriyanov, A. Tashlinsky, and V. Dementiev. "*Detailed Clustering Based on Gaussian Mixture Models*", In: Arai, K., Kapoor, S., Bhatia, R. (eds) Intelligent Systems and Applications. IntelliSys 2020. Advances in Intelligent Systems and Computing, Springer, Cham, 1251, pp 437–448, (2020),

[3] G. McLachlan, S. Rathnayke. "*On the number of components in a gaussian mixture model*", Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 4(5), pp. 341–355 (2014).

[4] Y. Zhang, M. Li, S. Wang, S. Dai, L. Luo, L. Zhu, H. Xu, X. Zh, C. Yao, and H. Zhou. "*Gaussian mixture model clustering with incomplete data*". ACM Transactions on Multimedia Computing, Communications, and Applications, *17*(1s), pp. 1–14, (2021).

[5] Y. Zhao, A. K. Shrivastava and K. L. Tsui, *"Regularized Gaussian Mixture Model for High-Dimensional Clustering*", *in IEEE Transactions on Cybernetics*, 49, pp. 3677-3688, (2019).

[6] J. Liu, D. Cai, and X. He. *"Gaussian Mixture Model With Local Consistency"*. Proceedings of the AAAI Conference on Artificial Intelligence, 24, pp. 512-7, (2010).

[7] M. Selwyn, and D. Difranco. "*The Application of Large Gaussian Mixed Models to the Analysis of 24 Hour Ambulatory Blood Pressure Monitoring Data in Clinical Trials*." Statistics in Medicine, 12, pp. 1665–1682, (1993).

[8] M. Yusoff, I. Mohamed and M. Bakar, *"Fraud detection in telecommunication industry using Gaussian mixed model",* 2013 International Conference on Research and Innovation in Information Systems (ICRIIS), pp. 27-32, (2013).

[9] Y. Wei, X. Yueru, F. Zhou, X. Shi, and Z. Ye. "*Investigation of Bus Drivers' Reaction to ADAS Warning System: Application of the Gaussian Mixed Model*" Sustainability 13, no. 16: 8759, (2021).

[10] F. Zorriassatine, A. Al-Habaibeh, R. Parkin, M. Jackson, and J. Coy, "*Novelty detection for practical pattern recognition in condition monitoring of multivariate processes: a case study*." Int J Adv Manuf Technol 25, pp. 954–963, (2005).

[11] V. Chauhan, B. Surgenor, *"Fault detection and classification in automated assembly machines using machine vision."* Int J Adv Manuf Technol 90**,** pp. 2491–2512 (2017).

[12] J. Yu, "*Machine Tool Condition Monitoring Based on an Adaptive Gaussian Mixture Model." ASME.* J. Manuf. Sci. Eng*,* 134(3), (2012)

[13] L. Qiu, S. Yuan, H. Mei, and F. Fang., *"An Improved Gaussian Mixture Model for Damage Propagation Monitoring of an Aircraft Wing Spar under Changing Structural Boundary Conditions"* Sensors 16, no. 3: 291. (2016)

[14] X. Shen, Y. Zhang, K. Sata and T. Shen, "*Gaussian Mixture Model Clustering-Based Knock Threshold Learning in Automotive Engines*," IEEE/ASME Transactions on Mechatronics, 25, no. 6, pp. 2981-2991, (2020).

[15] M. Yusoff, C. Ooi, M. Lim, M. Leong, "*A Hybrid K-Means-GMM Machine Learning Technique for Turbomachinery Condition Monitoring."* MATEC Web of Conferences, 255, (2019).

[16] T. Heyns, P.S. Heyns, J.P. de Villiers, *"Combining synchronous averaging with a Gaussian mixture model novelty detection scheme for vibration-based condition monitoring of a gearbox"* Mechanical Systems and Signal Processing, 32, pp. 200-215 (2012)

[17] T. Ayodele, *"Types of Machine Learning Algorithms*". New Advances in Machine Learning, edited by Yagang Zhang, IntechOpen, (2010).

[18] D. Holmes, and L. Jain. "*Innovations in Machine Learning : Theory and Applications"*. Springer, (2006).

[19] Kwedlo, W. "*A New Method for Random Initialization of the EM Algorithm for Multivariate Gaussian Mixture Learning*". In R. Burduk, K. Jackowski, M. Kurzyński, M. Woźniak, and A. Żołnierek (Eds.), Proceedings of the 8th International Conference on Computer Recognition Systems cores 2013 (Vol. 226). essay, Springer, (2013).

[20] O. Parsons, "*A Gaussian Mixture Model Approach to Classifying Response Types*". In N. Bouguila & W. Fan (Eds.), Mixture models and applications. essay, Springer, (2020).

[21] M. Hinz, M. Radetzky, L. H. Guenther, P. Fiur, S. Bracke, "*Machine Learning Driven Image Analysis of Fine Grinded Knife Blade Surface Topographies"* Procedia Manufacturing, 39, pp. 1817-1826, (2019).

[22] E. Sancaktar, R. Gomatam. "*A Study on the Effects of Surface Roughness on the Strength of Single Lap Joints."* Journal of Adhesion Science and Technology, 15, pp. 97-117. (2001)

[23] C. Fan, M. Chen, X. Wang, J. Wang, and B. Huang, "A review on data preprocessing techniques toward efficient and reliable knowledge discovery from building operational data. " Frontiers in Energy Research, 9, (2021).

[24] L. Shalabi, Z. Shaaban and B. Kasasbeh, "Data Mining: A Preprocessing Engine", J. Comput. Sci., 2: 735-739, (2006).

[25] N. Shimizu, & H. Kaneko. "Direct inverse analysis based on Gaussian mixture regression for multiple objective variables in material design." Materials & Design. (2020)