A Novel Perspective towards SVM Combined with Autoencoder

Deqiang Zou[1], **Hongtao Man**[2]

[1,2] Inspur Artificial Intelligence Research Institute, Shangdi 7 Street, Beijing, China

E-mail: zoudeqiang01@inspur.com

Abstract. In this paper, we propose a novel perspective towards the hybrid algorithm about support vector machine combined with neural network. We suggest that the depth of convolution neural network is supposed to insight the view of machines to acquiring an equal level of features as human do. The kernel function of support vector machine can be grasped flexibly where the neural network makes an efficient cross calculation for features exactly instead of the kernel function but more adjustable. To develop such a coincident format, we build a hybrid model with the half former part of autoencoder working as the kernel function and support vector machine working as the core classifier, with certain ways to train the hybrid model: discrete, continuous and prejudice. The hybrid model inherits asset of each algorithm, and that process is generally subject to the objective perspective. We take the hybrid model to Covid 19 detection compared with other well-performed models, and experimental results illustrate that our perspective is advisable which achieves a state-of-the-art performance in medical scheme.

1. Introduction

Hybrid model techniques have a long record of showing better performance in kinds of applications, which derive from ”no free lunch” drawback in machine learning[1]. The hybrid models extract advantages from individual models and then exert a superposed effect in the applications to adapt to complex scenes. The mechanism of the hybrid models depend on how to look upon the algorithms that possess particular characteristics and make logical composition, generally expecting a great effect.

Machine learning is a subfield of AI study and deep learning is the popular branch of machine learning[2]. If compared in characteristics, machine learning can fit a small sampling learning and owns a complete theory, but machine learning is limited to the ability and even needs subjective participation for adjustment[3]. On the contrary, deep learning relies on massive data and exists a ”black box” effect that is never expected in interpretation, but deep learning is powerful and intelligent enough[4]. Currently, the latter approach is far more frequently-used due to the state of the explosion of data and the increment of hardware performance[5]. Deep learning works well and displays a huge potential in a series of intelligent scenes[6]. Nonetheless not always in some domains like medical, financial, and precision industries, they are generally supported with high accuracy, stability, and interpretability[7, 8]. It is still troublesome to develop a model to adapt to the above requirements and make sense in certain applications.

To accommodate for certain demands in applications, the hybrid models are advisable schemes that refine multiple algorithms with the deduction of mathematics theory, and then take the positive patches to construct an objective-function model. The new model is supposed to inherit the positive characteristics of parent algorithms and adapt to the practical scenes no matter of accuracy, stability, even interpretability, etc, and stands out among unit algorithm[9]. The Norway specialists construct 12 hybrid models to predict the demand for endocrinology, gastroenterology, vascular, urology, and pediatric surgical units, involving seasonal autoregressive integrated moving average, support vector regression, multilayer perceptron, and long short-term memory[10]. The USA economists present a cluster based classification model, comprises of K-means clustering and a fitness-scaling chaotic genetic ant colony algorithm, which performs more robustness in financial crisis prediction[11]. Industry 4.0 describes the combination of machine learning and simulation to somewhat promote the development of hybrid system[12]. Moreover, the hybrid model is not only a mere composition of algorithms in techniques but also deserves a novel perspective on comprehension of all kinds of algorithms[13]. The predominant effect of hybrid models is intended to be a problem-oriented scheme in practical scenes[14]. Generally, a novel perspective towards algorithms goes ahead of the hybrid model.

2. A novel perspective towards hybrid algorithm

2.1. Convolution depth of the image

Data-centric AI is attractive now and expects to pay much more attention to data sources rather than models[15, 16]. Even though a small sampling could train a well-performed model, a huge network is not necessary. Data sources are the basement of the upper performance in the project. If we ignored the data ingredients and pursue much on the models, it would lead to an upper limit of efficiency and get into a plight.

One challenge of deep learning is to figure out relations between features and the depth of neural networks resulting in reflecting the necessity of the depth. Although convolution neural networks and their variates perform well in processing image data, the ability of these algorithms to learn features is still far from what is observed in humans[17]. The relevant studies of quantifying the feature-depth correspondence present that shallow neural networks usually produce inefficient results and deep neural networks are not necessarily behaving well either[18]. The convolution depth is of concern to the image data, and we expect the insights of both humans and machines to be kept at an equal level, which is convenient for value consistency. The convolution depth to some extent is the evolution degree of image data, which reflects the features transformation of image data in the convolution process. We expect the machine just acquire a real object rather than extremely abstract knowledge which never occurred in the real world. At least, the principle of the convolution process is supposed to avoid image data value decreased, so machines can learn complete image data for tasks. It is certainly worth establishing such a principle.

Considering a lot about the principle of convolution process, autoencoder is an efficient way to properly learn the implication of image data and represent image data as a whole, which transformation is regarded as a low image data decreased[19]. The image data representation is necessary for a success of applications in computer version and even occupies a decisive position in cognition science[20]. An autoencoder aims at setting the target values to be equal to the inputs, that middle layers reserve image data as much as possible in transformation process[21]. It consists of two parts of the encoder and the decoder, which are defined as transitions ϕ and ψ, such that

$$\begin{cases} \phi : \mathcal{X} \rightarrow \mathcal{F}, \\ \psi : \mathcal{F} \rightarrow \mathcal{X}, \\ \underset{\phi,\psi}{\arg\min} \, ||\mathcal{X} - (\phi \circ \psi)\mathcal{X}||^2. \end{cases} \tag{1}$$

Generally, the feature space $\mathcal{F} \in \mathbb{R}^q$ is lower dimension than the input space $\mathcal{X} \in \mathbb{R}^p$, $q \leq p$. Assume the latent image h in space $\mathcal{F}$, and the latent image h can be regarded as a compressed representation of input image data. Set a symmetric loss L,

$$L = \frac{||\mathcal{X} - (\phi \circ \psi)\mathcal{X}||^2}{2}. \tag{2}$$

For a convolution neural network, the max pooling process[22] generally takes away the inconspicuous pixels of the image which may cause a loss to the image data that is unexpected in our work. Therefore, we decide to take the mean pooling process instead, or even no pooling process in the context hybrid model.

2.2. Kernel stage of SVM

SVM(Support Vector Machine) is a well-known method of machine learning[23], which core target is that find a hyper plane $f(\mathbf{w}) = \mathbf{w}^T\mathbf{x} + b$ to maximal split the two clusters of points: Assume that data $\{(\mathbf{x_i}, y_i)\}_{i=1}^N, \mathbf{x_i} \in \mathbb{R}^p, y_i \in \{+1, -1\}$, and the optimization function is

$$\begin{cases} \min\limits_{\mathbf{w},b} \frac{1}{2}\mathbf{w}^T\mathbf{w}, \\ s.t. \quad y_i(\mathbf{w}^T\mathbf{x_i} + b) \geq 1, \forall i = 1, 2, ..., N, \end{cases} \tag{3}$$

introduce the penalty,

$$\begin{cases} \min\limits_{\mathbf{w},b} \frac{1}{2}\mathbf{w}^T\mathbf{w} + C \cdot \sum\limits_{i=1}^{N} \xi_i, \\ s.t. \quad y_i(\mathbf{w}^T\mathbf{x_i} + b) \geq 1 - \xi_i, \forall i = 1, 2, ..., N, \xi_i \geq 0, \end{cases} \tag{4}$$

$\xi_i = 1 - y_i(\mathbf{w}^T x_i + b)$ is the slack variable and C is a coefficient of the penalty to make a trade off between the maximization of the margin and the minimization of the classification errors. It is a 'QP' quadratic programming problem corresponding to a certain solution[24]. But generally, the optimization problem is usually solved with the help of dual problem which is created out by Lagrange formulation, and after a series of deduction the optimization function is

$$\begin{cases} \min\limits_{\lambda} \frac{1}{2} \sum\limits_{i=1}^{N} \sum\limits_{j=1}^{N} \lambda_i\lambda_j y_i y_j \mathbf{x_i}^T\mathbf{x_j} - \sum\limits_{i=1}^{N} \lambda_i, \\ s.t. \quad \lambda_i \geq 0, \lambda_j \geq 0, \sum\limits_{i=1}^{N} \lambda_i y_i = 0, \end{cases} \tag{5}$$

of which λ_i is the hyper-parameter of Lagrange formulation. In above Eq.5, the vector product $\mathbf{x_i}^T\mathbf{x_j}$ is an implement which used to be adorned for non-linear classification — kernel trick, $\kappa(\mathbf{x_i}^T, \mathbf{x_j})$ makes effect that calculate a dot product in low dimension corresponding to high dimension with no need of space mapping but also achieve the target. The kernel trick $\kappa(\mathbf{x_i}^T, \mathbf{x_j})$ is a highlight for SVM which brings a leading role in image classification in machine learning, and commonly used kernel functions are sigmoid kernel, radial basis function, etc. Refer to relevant researches for more details about the kernel functions[25].

A novel perspective that the kernels of SVM are adaptive forms which can be generally imitated by neural network. The neural network can automatically study and conduct the features that $\kappa(\mathbf{x_i}^T, \mathbf{x_j})$ can be fitted with the optimization process of algorithms, which just works as a flexible kernel of SVM. The flexible kernel is adaptive to SVM which transforms the features into a separable level, and then make an efficient classification. Generally, the flexible kernel functions are achieved with backward propagation using gradients.

Although the system of SVM-relative hybrid algorithms have been developed greatly with commonly used convolution neural networks in computer version[26, 27, 28], the system just considers that convolution neural networks extract features in image data and take the so-called features to SVM without constraints, which misleads the effect of each part. This perspective of the system usually produces a model with weak robustness because the standard is submerged. Thus some papers are written to demonstrate its feasibility because the results of the system are not stable, even less than common neural networks which based on softmax layer as final decision[29]. Generally, the deployment of SVM-relative hybrid algorithms can be designed to discrete way and continuous way and it depends on the training way. The discrete way is that separately train the parameters of each part of the hybrid model ,and then take the intermediate data of neural networks to SVM. The continuous way is that entirely train the parameters of the hybrid model in somewhat design with the help of mathematics theory, and the optimization of each part is simultaneous.

3. The Model

To develop such a coincident format in Sec.2, we propose a hybrid model in that half former part of the autoencoder works as kernel function and SVM works as the core classifier, structure as Fig.1. Not only does autoencoder extremely decrease the image data loss in the convolution process, but also makes anomaly detection with its model characteristic in the meanwhile[30], which gradually detaches the image data with their features if image data come from different cases. Especially combined with the state of image data gradually detached, corresponding to the characteristic of SVM using a hyperplane maximum to split data points, it is easier to find an optimal hyperplane and fully exerts the power of SVM.

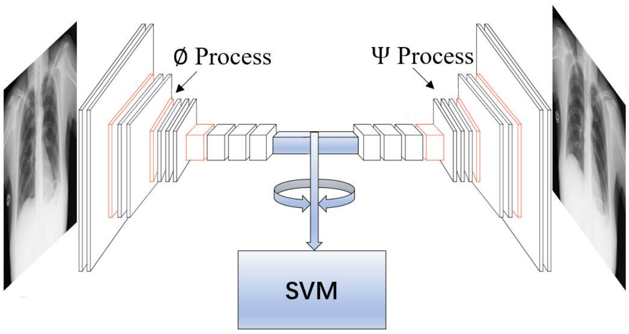


Figure 1. SVM combined with autoencoder.

In subsequent experiments, we would take 3 diverse ways to train the hybrid model of SVM combined with autoencoder, and each way is highly expected. Thus, the 3 training ways as follows:

(a) Discrete way to train the hybrid model

Use both positive case and negative case to train an autoencoder model getting the encoder part, and then take the latent data into SVM classifier as input, thus separately train the hybrid model;

(b) Continuous way to train the hybrid model

Use both positive case and negative case to train an autoencoder model getting the encoder part, and then take the former part of the autoencoder as initializer contacting with SVM, thus continuously train the hybrid model. To deal with the continuity, we intend to differentiate SVM object with respect to the activation of the latent layer. Let the objective in Eq.4 be $p(\mathbf{w})$, and the input $\mathbf{x}$ is replaced with the latent activation $\mathbf{q}$, then

$$\frac{\partial p(\mathbf{w})}{\partial \mathbf{q_i}} = -Cy_i\mathbf{w}(\mathbb{I}\{1 \geq y_i\mathbf{w}^T\mathbf{q_i} + b\}), \tag{6}$$

and for L2-SVM, we have

$$\frac{\partial p(\mathbf{w})}{\partial \mathbf{q_i}} = -2Cy_i\mathbf{w}(\max(1 - y_i(\mathbf{w}^T\mathbf{q_i} + b), 0)), \tag{7}$$

once you've made that, backpropagation algorithm exactly performs on the hybrid model;

(c) Prejudice way to train the hybrid model

Autoencoder possesses the attribute itself to make anomaly detection which gradually detaches the data with features. We just randomly choose one case image data to train an autoencoder getting the encoder part and then transform both positive and negative cases into latent data with this encoder, eventually taking these 2 cases of latent data into the SVM classifier to train the hybrid model.

4. Experiment

Computer-aided diagnosis is a rapidly growing dynamic area of research[31, 32]. Covid 19 pandemic affects the healthcare delivery systems and makes a far reaching life for people across the world. And a large number of studies have focused on Covid 19 with AI models which provide an auxiliary diagnosis tool. With the studies for diagnosing Covid 19, SVM is the most widely used machine learning mechanism, and convolution neural network is the most widely used deep learning mechanism[33].

We collect the Covid 19 datasets from Kaggle which contains 500 chest X-ray images of the normal case and covid case all over the world[34]. And then, we make a series of processes to dispose of these images, including splitting the dataset into train group, validation group and test group, resizing images, label encoding, data augmentation, etc. We find the proportion of chest in the image sensitively impacts the evaluations of the hybrid model that the tiny segmentation of the image is much desired. We process the chest X-ray images as Fig.2

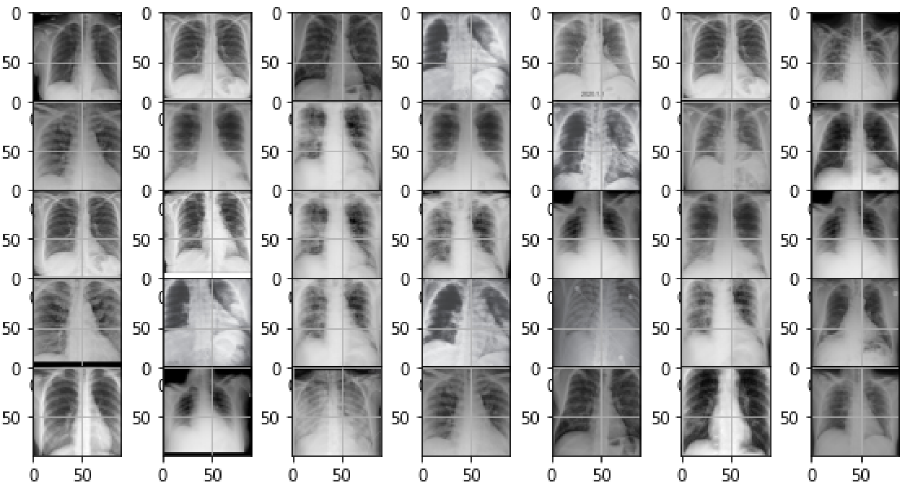


Figure 2. The chest X-ray images of Covid 19 patients.

We train the autoencoder to get the latent data which can reserve image information as much as possible. Shown in the middle of Fig.3, it is the latent image of a covid case that gives a blurry format. We calculate the image loss with Eq.2 using mean square error in latent space with 25 epochs, that the convolution process just makes a low image data decreased that is closed to 0.001.

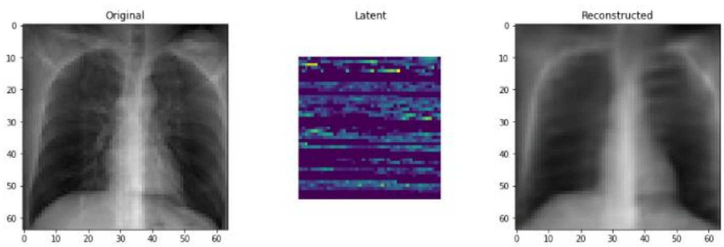
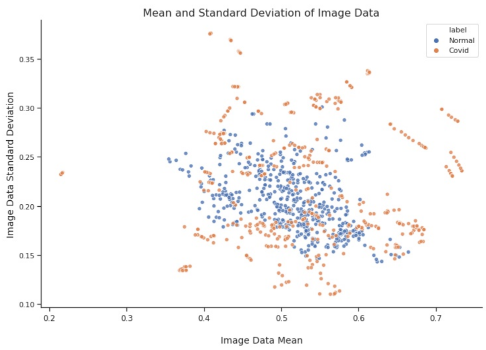
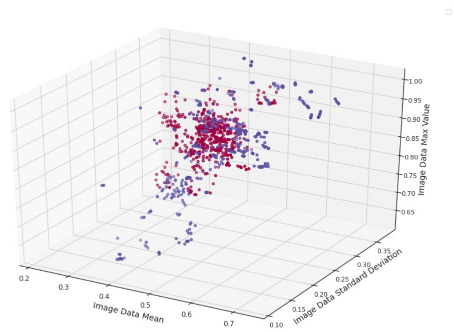


Figure 3. The latent image of autoencoder.

We initially explore the chest X-ray images with the features of images that plot scatters with mean values and standard deviations showing the distribution of data points. In Fig.4(a), we find it hard to directly separate these data points with the SVM model. Thus we take autoencoder to transform the image data to latent images data, and in Fig.4(c), plot the same pattern of the latent image data which presents a clear margin between the normal case and covid case. This phenomenon is much identical to the kernel function of SVM and proves that autoencoder has fitted a flexible kernel function. And this result gives it a possibility to split the chest X-ray image data with SVM. Although the 2-dimension scatters present a great effect, the higher dimension scatters are expected better, 3-dimension scatters in Fig.4(b) and Fig.4(d).

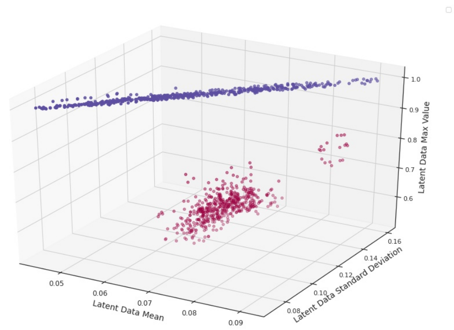


(a) 2-dimension of image data



(b) 3-dimension of image data

(c) 2-dimension of latent data



(d) 3-dimension of latent data

Figure 4. Data points distribution of image data and latent image data with scheme (c).

Table 1. Performance of the models in evaluation.

	Accuracy	Sensitivity	F1-score
Vgg	0.963	0.921	0.94
$Resnet$	0.965	0.953	0.959
$Autoencoder - SVM^{(a)}$	0.971	0.968	0.945
$Autoencoder - SVM^{(b)}$	0.975	0.988	0.943
$Autoencoder - SVM^{(c)}$	0.997	1.00	1.00

To keep the robustness of models, we take the models to several Covid 19 datasets and make a synthetic consideration. And we know that sensitivity is an important factor that false negatives can bring a terrible consequence for Covid 19 detection. The hybrid system of SVM performs well on the whole and $Autoencoder - SVM^{(c)}$ which is trained with bias has an accuracy of 0.996 and a sensitivity of 1.00. We find only 1 case is misjudged in the confusion matrix of $Autoencoder - SVM^{(c)}$. Besides that, we plot the roc curve of each model in Fig.5 and the hybrid model $Autoencoder - SVM$ system shows a good effect which AUC is close to 1.

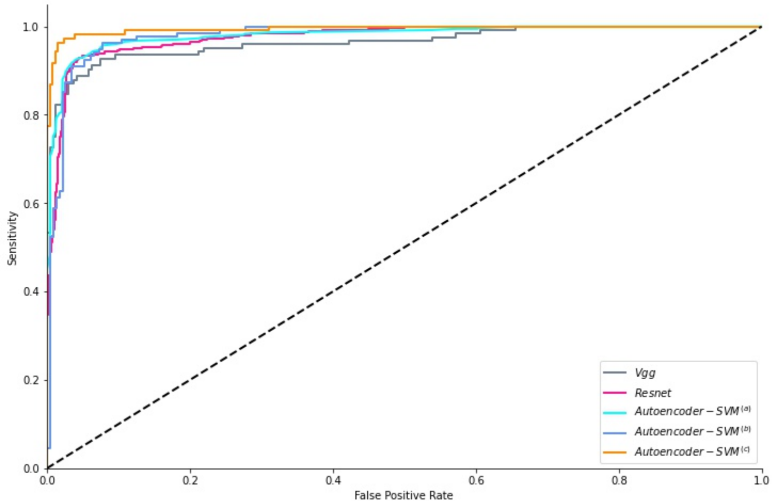


Figure 5. Roc curves of the models.

In this comparison part, we review the modern techniques in Covid 19 diagnose that most techniques focus on the architecture of neural networks, in which popular base choices are Vgg and Resnet, that results are generally presented in Table.1. Considered more about hybrid models, each of them requires a set of specific environments, and it is unfair to make a comparison among the different configurations, especially for a complex algorithm. Besides that, we intend to perform several hybrid models, such as enhanced and adaptive-genetic algorithm-multilayer perceptron[35], Vgg and slap swarm algorithm[36], advanced convolution neural network with grasshopper optimization algorithm[37] and etc, yet they just perform an ordinary level, at least no closed to 1.00.

In the context, we choose 3 various levels of severity covid cases and take the saliency map to explain the hybrid model $Autoencoder - SVM$. We calculate gradients of each pixel to outputs in 3 images and then plot saliency images with the gradients in Fig.6. And according to medical knowledge, it gives a professional clue with highlight pixels for diagnosing Covid 19.

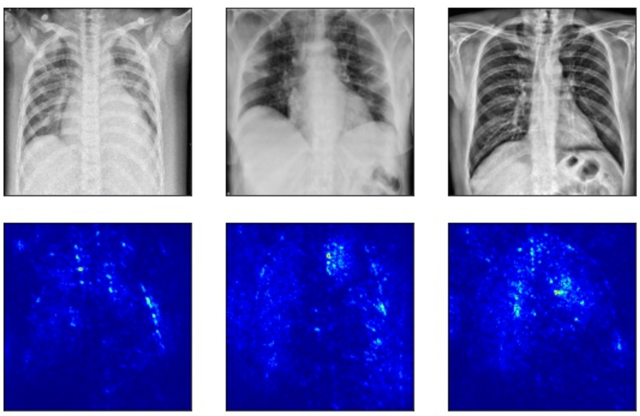


Figure 6. Saliency map with $Autoencoder - SVM$ on Covid 19.

Remark 1 *In our experiments, we find it is too slow to train an autoencoder without the pooling process because it generally required a huge computing resource. Thus it doesn't matter to take the pooling process into the hybrid models in practical. Whereas the no-max-pooling deal is deserved when the states are the same number of parameters.*

5. Conclusion

This paper presents a novel perspective in 2 aspects. The convolution depth is of great significance in computer version, to some extent reflecting the features transformation of image data, and we expect the insight of machines to keep at an equal level with human, at least avoiding image data value decreased in the convolution process. The kernels of SVM are flexible forms which can be adaptively learned with the backward propagation process of deep neural networks, that brings a proper entrance to develop the performance of SVM. Thus, we proposed a hybrid model $Autoencoder - SVM$ conformed with the above perspective(or principle), and then take $Autoencoder - SVM$ to Covid 19 detection compared with other well-performed models, in which presents a great effect. Covid 19 detection attributes to the medical field which is endowed with limitations in some aspects. The hybrid model $Autoencoder - SVM$ is applied to a 500 Chest X-Ray images dataset without over-fitting states which is a common phenomenon in small-sampling medical scenarios, and it is supported with a brilliant theory though the saliency maps can be referred for interpretability in medical AI.

References

[1] Ardabili S, Mosavi A and Várkonyi-Kóczy A 2020 *Advances in Machine Learning Modeling Reviewing Hybrid and Ensemble Methods* (Engineering for Sustainable Future)
[2] Shinde P P and Shah S 2018 A review of machine learning and deep learning applications *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)* pp 1–6
[3] Dogan A and Birant D 2021 *Expert Systems with Applications* **166** 114060 ISSN 0957-4174
[4] Buhrmester V, Münch D and Arens M 2021 *Machine Learning and Knowledge Extraction* **3** 966–989 ISSN 2504-4990
[5] Zhang L, Tan J, Han D and Zhu H 2017 *Drug discovery today* **22** 1680–1685
[6] Alzubaidi L, Zhang J, Humaidi A J, Al-Dujaili A, Duan Y, Al-Shamma O, Santamaría J, Fadhel M A, Al-Amidie M and Farhan L 2021 *Journal of big Data* **8** 1–74
[7] Garg A and Mago V 2021 *Computer Science Review* **40** 100370
[8] Nguyen B D, Do T T, Nguyen B X, Do T, Tjiputra E and Tran Q D 2019 Overcoming data limitation in medical visual question answering *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Springer) pp 522–530
[9] Nalisnick E, Matsukawa A, Teh Y W, Gorur D and Lakshminarayanan B 2019 Hybrid models with deep and invertible features *International Conference on Machine Learning* (PMLR) pp 4723–4732
[10] Aravazhi A 2021 *AI* **2** 512–526 ISSN 2673-2688
[11] Uthayakumar J, Metawa N, Shankar K and Lakshmanaprabu S 2020 *Information Systems and e-Business Management* **18** 617–645
[12] Von Rueden L, Mayer S, Sifa R, Bauckhage C and Garcke J 2020 Combining machine learning and simulation to a hybrid modelling approach: Current and future directions *International Symposium on Intelligent Data Analysis* (Springer) pp 548–560
[13] Khan F N, Fan Q, Lu C and Lau A P T 2019 *Journal of Lightwave Technology* **37** 493–516
[14] Paleyes A, Urma R and Lawrence N D 2020 *CoRR* **abs/2011.09926**
[15] Hajij M, Zamzmi G, Ramamurthy K N and Saenz A G 2021 *arXiv preprint arXiv:2110.02491*
[16] Motamedi M, Sakharnykh N and Kaldewey T 2021 *arXiv preprint arXiv:2110.03613*
[17] LeCun Y 2012 Learning invariant feature hierarchies *European conference on computer vision* (Springer) pp 496–505
[18] Han Z, Yu S, Lin S B and Zhou D X 2020
[19] Louizos C, Swersky K, Li Y, Welling M and Zemel R 2015 *arXiv preprint arXiv:1511.00830*
[20] Lingheng M, Ding S, Zhang N and Zhang J 2018 *Neural Computing and Applications* **30**
[21] Wang W, Huang Y, Wang Y and Wang L 2014 Generalized autoencoder: A neural network framework for dimensionality reduction *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* pp 490–497
[22] Christlein V, Spranger L, Seuret M, Nicolaou A, Král P and Maier A 2019 Deep generalized max pooling *2019 International Conference on Document Analysis and Recognition (ICDAR)* pp 1090–1096
[23] Suthaharan S 2016 Support vector machine *Machine learning models and algorithms for big data classification* (Springer) pp 207–235

[24] Gill P E and Wong E 2015 *Mathematical programming computation* **7** 71–112
[25] Murty M N and Raghava R 2016 *Kernel-based SVM* (Support Vector Machines and Perceptrons)
[26] Eid M M and Elawady Y H 2021 *European Journal of Electrical Engineering and Computer Science* **5** 1–8
[27] Çınar A and Tuncer S A 2021 *SN Applied Sciences* **3** 1–11
[28] Kilicarslan S, Adem K and Celik M 2020 *Medical Hypotheses* **137** 109577 ISSN 0306-9877
[29] Agarap A F 2017 *CoRR* **abs/1712.03541**
[30] Chen Z, Yeo C K, Lee B S and Lau C T 2018 Autoencoder-based network anomaly detection *2018 Wireless Telecommunications Symposium (WTS)* (IEEE) pp 1–5
[31] Zhang J, Li K, Zheng L, Zhang J, Ren Z, Song T, Yu H, Yang Z, Wang L and Jiang L 2020 *Frontiers in Cellular and Infection Microbiology* **10** ISSN 2235-2988
[32] Takahashi R and Kajikawa Y 2017 *International Journal of Medical Informatics* **101** 58–67 ISSN 1386-5056
[33] Alyasseri Z A A, Al-Betar M A, Doush I A, Awadallah M A, Abasi A K, Makhadmeh S N, Alomari O A, Abdulkareem K H, Adam A, Damasevicius R *et al.* 2021 *Expert systems* e12759
[34] Tahir A M, Chowdhury M E, Khandakar A, Rahman T, Qiblawey Y, Khurshid U, Kiranyaz S, Ibtehaz N, Rahman M S, Al-Maadeed S, Mahmud S, Ezeddin M, Hameed K and Hamid T 2021 *Computers in Biology and Medicine* **139** 105002 ISSN 0010-4825
[35] Mishra S, Tripathy H K, Mallick P K, Bhoi A K and Barsocchi P 2020 *Sensors* **20** ISSN 1424-8220
[36] Sahlol A T, Kollmannsberger P and Ewees A A 2020 *Scientific Reports* **10** 1–11
[37] Tyagi A and Mehra R 2021 *SN Applied Sciences* **3** 1–14