

Review

Regularization-Based Solutions to Overfitting in Modern Predictive Models: A Review

Osita Miracle Nwakeze^{1,*}

¹ Department of Computer Science, Chukwuemeka Odumegwu Ojukwu University, Uli, Anambra State, Nigeria;
ma.nwakeze@coou.edu.ng

* Correspondence

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Abstract: One of the most serious issues in the current predictive modelling is overfitting, especially with the increase in the complexity and size of machine learning systems. This research proposes regularisation-based methods as critical methods of enhancing generalisation and avoiding noise memorization in training data by models. The systematic literature review of the study that includes 2006-2025 research provides the synthesis of classical regularization methods, including L1/L2 penalties, Elastic Net, dropout, and early stopping, and emerging methods, including probabilistic dropout variants, Bayesian regularization, adaptive regularizers, and hybrid frameworks. The review points to the importance of the regularisation in the context of increasing the performance of generalisation, enhancing robustness to noisy or finite-sized datasets, stabilising optimization dynamics, and interpretability in high dimensional computations. It also determines the major shortcomings in the extant research such as, lack of comprehension on implicit regularisation, the cross domain comparative assessment and the requirement of adaptive and automatic strategies of regularisation. The paper ends with the research recommendations and open directions of research that are meant to enhance the theory, diagnostic tools, and towards the practitioners to effective regularisation configurations under different data regimes. Altogether, this paper gives an integrative and holistic approach to regularisation as a core building block of constructing credible, robust and general predictive models.

Keywords: Overfitting; Regularization; L1/L2 Penalties Dropout; Machine Learning; Generalization.

Copyright: © 2025 by the authors. This is an open-access article under the CC-BY-SA license.



1. Introduction

The issue of overfitting, i.e. the ability of a predictive model to capture noise or domain specifics in the training data instead of the underlying signal, has been a primary challenge to effective model use in new domains (vision, healthcare, finance). Modern models, with their growing capacity in a dramatically increasing capacity, can perfectly interpolate training data, but will also demonstrate a variable generalisation performance in practise, which has fuelled new theoretical and empirical inquiry into the causes and limits of generalisation [1], [2]. Awareness of the trade-offs between model expressivity and generalisation is hence a conditioning factor in the development of principled mitigation procedures when executing model on real-world, and potentially noisy, datasets [3].

Regularisation - the general concept of improving the generalisation performance of a model to unseen da-

ta, in other words, all the techniques and procedures that help to enhance the generalisation ability of a model which comprises of a wide range of methods that include classical norm penalties (L1/L2) as well as more recent stochastic and algorithmic ones, including dropout, data augmentation, batch normalisation, and early stopping [2], [4]. Experimental evidence has demonstrated that certain techniques are highly effective in certain contexts (such as dropout in deep networks and use of explicit penalty terms in gradient-boosted trees) and other studies emphasize that optimization dynamics and regularization implicit in training algorithms are also significant contributors to generalization performance [5], [6]. These results combined proposed that (1) to protect overfitting in current predictive pipelines, a hybrid approach of data-level, architecture-level, and algorithm-level protections is expected to be necessary; and (2) strategies to

eliminate overfitting in current predictive pipelines are ineffective without addressing all three levels of control.

Although the set of techniques is quite extensive, there are also significant gaps: comparative assessments between tasks and data regimes are still at the stage of development, the relations between explicit and implicit regularizers have not been fully described, and guidelines of the best practices are not yet fully formed in terms of choosing and tuning regularization combinations [2], [4]. The proposed study attempts to fill these gaps through survey-based research on regularisation-based solutions, synthesis of empirical evidence on various areas of problems, and offer practical advice to researchers and practitioners who have to trade off model capacity, robustness, and interpretability during real world deployments. Combining the recent theoretical findings and empirical outcomes, the study will have the goal of generating practical recommendations on how to decrease overfitting and maintain predictive power at the same time.

The rest of this paper is organized as follows: Section 2 reviews related works on regularization techniques for mitigating overfitting. Section 3 details the systematic literature review methodology employed. Section 4 explains the core concepts of overfitting and regularization, including their advantages in modern predictive models. Section 5 concludes the paper with key findings and discusses open research directions.

2. Related Works

The researchers across the reviewed literature proposed various regularisation approaches to reduce overfitting and enhance the model generalisation in different fields of machine learning. [7] developed a L1/4 regularisation scheme with S-inclined activation function, which with improves sparsity, decreases prediction errors and demands less samples than L1 and L1/2. [8] showed that Lasso regularisation enhances feature selection and prediction accuracy of air quality prediction especially of particulate matter. [9] solved the overfitting problem in kernel learning models by proposing Bayesian hyperparameter regularisation, which greatly enhanced the performance of LS-SVM in the situation that contained numerous kernel parameters. [2] outlined a broad overview of conventional and deep-learning regularisation measures, their strengths, weaknesses, and the necessity of their further theoretical insights. [10] built upon CNN regularization and introduced the Probabilistic Feature Importance Dropout (PFID), which addresses the dynamically changing dropout rate and demonstrated significant improvements in the accuracy of CIFAR-10, MNIST, and Fashion-MNIST.

A few works were devoted to the causes of overfitting and the focus on practical mitigation strategies. [11] described the causes of overfitting and noise, data scarci-

ty, and model complexity, and mentioned early stopping, network reduction, data expansion, and regularisation as the fundamental solutions. According to [12], the Regularised Least Squares Ensemble Model (RLS-EM) was proposed, which enhances the accuracy and stability of a high-dimensional engineering design problem because regularisation and sample augmentation are combined. As demonstrated by [13], optimizer choice is interacted with regularisation whereby NAG, Adam, AdaMax and Batch Normalisation enhance the CNN generalisation. In the L1-based feature selection of a big-data precipitation forecasting model, [14] show that ensemble classifiers have a significant accuracy improvement when using datasets with up to 10 million samples. [15] also outlined some general progress in regularization in various high-dimensional ML tasks with various contributions in a curated special issue.

Other researches proposed domain-specific or hybrid regularizations. [16] combined ensemble learning and CNNs in the ELVD model, achieving up to 99% accuracy in noisy traffic sign recognition. [17] improved Stochastic Configuration Networks by integrating Elastic Net regularization and dynamic loss coefficients, reducing overfitting by over 50% across benchmarks. [18] has associated regularisation with Bayesian inferences of inverse imaging issues and highlighted hybrid physics-based and data-driven schemes. [19] demonstrated L2 regularisation is better at low-dimensional ANN tasks compared to the L1 and dropout, and dropout variants are better when the ANN is high-dimensional, such as MNIST. According to [20], the feature reduction feature of autoencoders reduces overfitting in IDS models, and Decision Trees and KNN have high accuracies with reduced feature sets. By incorporating both Bayesian-regularised RNNs and NSGA-II, [21] obtained outstanding results in terms of space-camera reflector optimization. [22] proposed SynerGNet, and it was demonstrated that the combination of augmented data and regularization does not significantly harm predictive performance in various assessment measures even though test loss is growing. [23] developed MDWS, a dropout-weight attenuation algorithm that had the lowest error rates and reliable convergence in CNNs in image recognition. Last but not least, [24] divided the effective CNN regularisation techniques into the three groups: data augmentation, internal feature manipulations, and label transformations, and focused on their reproducibility and generalisation influence.

3. Research Methodology

Systematic Literature Review (SLR) methodology was taken in order to find, assess, and integrate academic evidence on regularisation-based measures to curb overfitting in contemporary predictive models. A systematic procedure shown in Figure 1 was used in this review

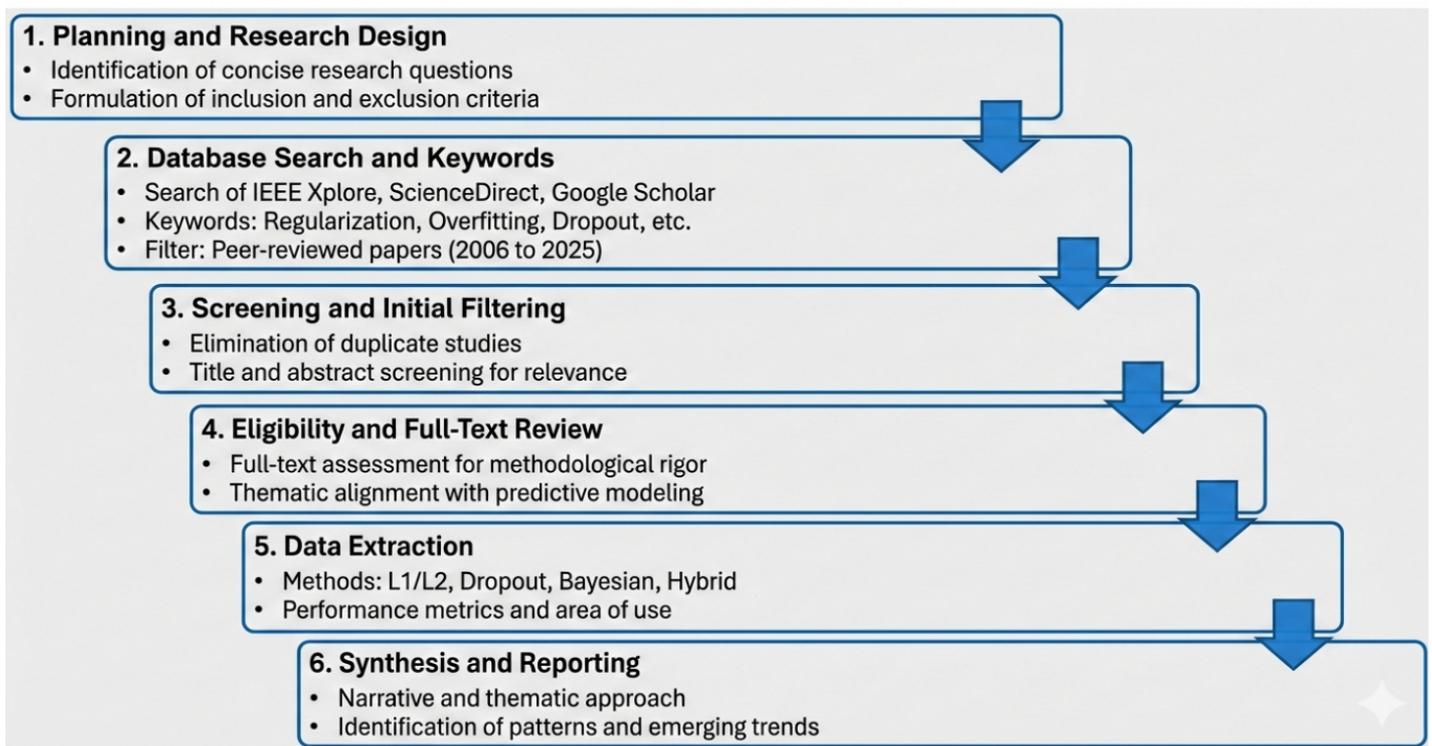


Figure 1. SLR stages.

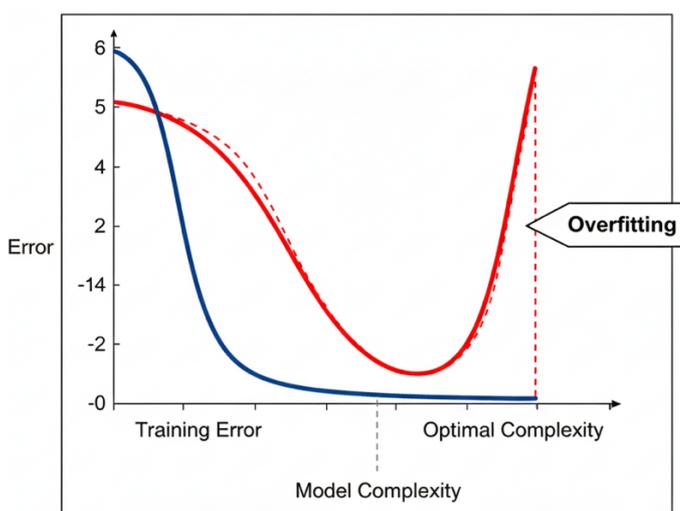


Figure 2. Example of Overfitting concept.

which included identification of concise research questions, the formulation of the inclusion and exclusion criteria and a search of major online academic databases, including IEEE Xplore, SpringerLink, ScienceDirect, ACM Digital Library and Google Scholar.

Such keywords as regularisation, overfitting, machine learning generalisation, dropout, L1/L2 penalties, implicit regularisation, and deep learning optimization were utilised to access the necessary peer-reviewed papers published in 2006 or later. Due to the elimination of duplicates, the studies were filtered by the title, abstracts, and full texts to identify relevance, methodological rigor, and relevant themes.

This SLR aims to provide a representative rather than exhaustive synthesis of regularization-based ap-

proaches to overfitting, focusing on influential peer-reviewed works published from 2006 onward (with inclusion extending to emerging preprints and accepted papers up to early 2025 to capture recent advancements). The selection considered here prioritized studies that demonstrates theoretical insights, empirical validation or novel methodological contributions, ensuring methodological rigor (clear experimental setups and benchmarks) and relevance to overfitting mitigation. While an exhaustive coverage of all publications was not feasible due to the rapid expansion of the field, the included works were chosen to achieve thematic balance across machine learning paradigms: classical methods (for instance norm-based penalties in linear models, SVMs, and ensembles like XGBoost) are represented alongside modern deep learning techniques such as dropout variants, probabilistic methods, and implicit regularization in large-scale neural networks. This balance reflects the evolution of regularization from explicit capacity control in smaller models to hybrid explicit-implicit strategies in overparameterized regimes, spanning applications in regression, classification, computer vision, time-series forecasting, and domain-specific tasks (e.g., healthcare, environmental modelling, and engineering optimization).

4. The Concept of Overfitting

In overfitting, it is not only the underlying relationship between the inputs and the targets that are being modelled by the learning algorithm but also the noise and idiosyncrasies of the training data, which results in a predictor that works well on the training data but fails to work well on the unseen data [25]. Classical statistical

explanations describe overfitting in the context of the bias variance trade-off: too flexible models may decrease bias, at the price of large variance, to be sensitive to sampling variability in the training data. Practically, overfitting is identified by the increasing discrepancy between training performance and validation/test performance: as training loss decreases and validation/test loss increases, then the model is presumably learning to memorise training-specific effects instead of learning generalizable structure [26]. Figure 2 illustrates an example of overfitting as a graph.

Cloudier high-capacity models (such as deep neural networks and large ensembles) confuse traditional intuitions as they can achieve near-zero training error on random labelled data, and nevertheless may often generalise well in practise revealing a vulnerability of the simple capacity elucidation of overfitting [1]. Such observation has led to finer insights: explicit regularizers (weight penalties), data-centric regularizers (augmentation, cross-validation) are effective in most contexts, still, optimization dynamics and implicit regularisation of algorithms (stochastic gradient descent) does play a significant role in determining whether a large model would overfit [1], [26]. Recent research on the so-called double-descent risk curve further indicates that at classical interpolation threshold and beyond, model capacity can sometimes lead to an error in the test, again highlighting that overfitting is a subtle phenomenon that involves model, data, and training interaction [27].

Since overfitting may occur in a variety of ways, including limited data, noisy labels, too complex architecture, or unsuitable optimization, it is normally multi-dimensional. Common techniques are norm-based penalties (L1/L2), stochastic regularizers using dropout and early stopping using validation monitoring as well as data augmentation in order to achieve lower model capacity, introduces noise, makes it harder to memorize and/or to represent the inputs better [26]. An effective experimental procedure integrates diagnostics (learning curves, hold-out validation), targeted regularization and hyperparameter search in such a way that model complexity is adjusted to the available data and the overall generalization demands of the task.

4.1. The Concept of Regularization

Regularization of machine learning is a collection of methods that allow overfitting to be avoided through limited model complexity. Overfitting happens when a model is fitted too well to the training data, it includes noise and irrelevant patterns which reduces its capacity to predict new data. Regularization is a method of solving this issue, which means to put constraints in the learning process, commonly with a penalty term in the loss term that is used to disfavour high or extreme values

of the parameters [5]. Regularisation techniques are usually divided into a number of methods. L2 regularisation (Ridge) includes a penalty that is proportional to the square of the model weights and thus that pushes the model weights towards smaller, smoother values [26]. L1 regularisation (Lasso) incorporates a penalty in relation to the absolute value of weights, which may reduce some weights to zero, which has the effect of performing feature selection. Dropout is a stochastic regularisation technique on neural networks in which, in training, randomly chosen neurons are disregarded, and the network is trained to learn the superfluous and more resilient features [5]. Early termination, data augmentation, and model architecture constraints are other measures and all geared towards enhancing generalisation.

Consistency is essential in the derivation of predictive models and this is more so when the data available are high dimensional or scarce. Regularization can stabilize, interpret, and be more resistant to noisy data by decreasing variance with a reasonable bias by censoring model complexity. The techniques are an essential part of modern deep learning and machine learning models to make sure that the model works on unseen data and not just remembers the training set [26].

4.2. Advantages of Regularization for Addressing Overfitting in Modern Predictive Models

Regularisation is a critical element in the contemporary machine learning due to its capacity to enhance the generalisation of the model, stability, and interpretability. The literature review has shown that various regularisation strategies including classical L1/L2 strategies to non-standard dropout-based strategies and Bayesian strategies all increase the reliability of predictive models across the various areas of application as shown in Figure 3. The subsequent subsections summarise the major perceived benefits that have been discovered in the reviewed literature.

4.2.1. Improved Generalization Performance

The ability of regularisation to enhance model generalisation through reducing over-complexity and avoiding the ability of models to capture noise in the training data is one of the most well-known advantages of regularisation. Algorithms like L1 and L2 and the Elastic Net, the large weights values are penalised, so that only the underlying structure patterns are being learned by the model. Regularisation of deep neural networks and regression-based models decreases prediction errors by a significant margin and the test-set accuracy as shown in empirical studies by [7] and [19]. The above findings demonstrate that regularisation promotes simpler, stronger models that are able to work reliably on unseen data points.

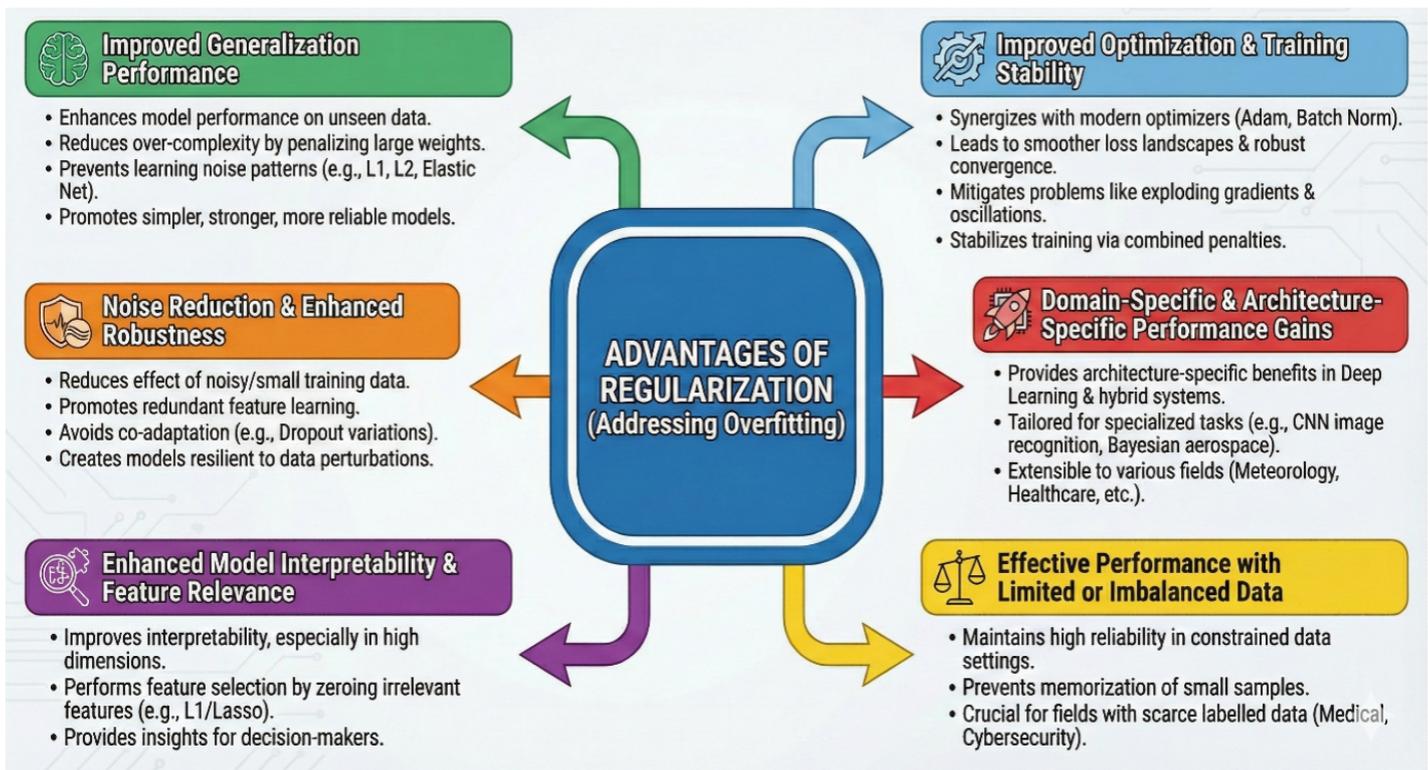


Figure 3. Advantages of Regularization for Addressing Overfitting.

4.2.2. Noise Reduction and Enhanced Robustness

The purpose of regularisation is important in reducing the effect of noisy or small training data, which is a frequent cause of overfitting. Such methods as dropout and its variations (e.g., Probabilistic Feature Importance Dropout by [10]) work by randomly taking away neurons in training. This promotes redundant feature learning as well as avoids co-adaptation between neurons and produces more resilient models to perturbations of data. [11] pointed out that regularisation of models to reduce their complexity is a good way of preventing overfitting caused by noise. These findings were validated in the reviewed works since the authors revealed that dropout-based strategies usually enhance robustness in image processing, time-series prediction, and sensor-based classification tasks.

4.2.3. Enhanced Model Interpretability and Feature Relevance

It is also regularised, which also gives it improved interpretability, especially in high-dimensional predictive modelling. The approaches based on L1 are particularly useful in terms of feature selection, reducing the importance of irrelevant features to a zero value. Researchers like [8] and [14] show that models using Lasso or elastic net regularisation can determine the most significant predictors to use in an environmental and meteorological forecasting exercise. This does not only give rise to easier models but also to insights that are useful to decision-makers that expect easy to understand outputs.

4.2.4. Improved Optimization and Training Stability

The other notable benefit of the reviewed studies is the enhancement of optimization behaviour. Regularization techniques play synergies with more recent optimizers and training procedures, like Adam, Nesterov momentum and batch normalization, resulting in more regular loss landscapes and more robust convergence. According to [13] and [17], regularised models exhibit fewer problems such as exploding gradients, oscillations, divergent parameters, etc. The regularization SCN architecture suggested by the Elastic Net that is presented by [17] showed in particular the way in which the combination of various regularization penalties allows stabilizing training and improving downstream performance.

4.2.5. Domain-Specific and Architecture-Specific Performance Gains

Another common point observed in the reviewed works is the ability of regularisation strategies to provide architecture-specific benefits in deep learning and hybrid systems. An example is the dropout with weight attenuation used to enhance the performance of image recognition in the dual-category CNN architecture suggested by [23], and Bayesian regularization to enhance the accuracy of reflector shape prediction in the aerospace engineering [21]. Such studies demonstrate that regularisation cannot be adopted as a universal method; it is beneficial, and it can be extended and applied in the specialised models of various areas of knowledge like meteorology, healthcare, transportation, and aerospace.

4.2.6. Effective Performance with Limited or Imbalanced Data

Regularisation becomes very important when the training data available is limited, skewed or hard to access. Regularization helps to have high reliability of performance even in constrained data setting by the fact that it does not make models to memorize small samples. Indicatively, [16] showed that regularised deep ensembles were highly recognised in noisy traffic-sign data, despite the shortage of data. The latter is especially beneficial in practise, when large labelled datasets are not always readily available, say in medical diagnostics, cybersecurity or remote sensing.

5. Conclusion

This paper looked at the regularisation-based remedies to the age-old problem of overfitting in current predictive models and summarises the findings of a broad spectrum of the literature that includes both classical machine learning and the current day deep learning. The review emphasised that overfitting occurs when models are fitted on the training data to the extent that they are able to emulate noise and spurious components within the training data which do not generalise to the unseen data, a phenomenon that is enhanced in high-capacity models such as deep neural networks. A systematic literature review has been used to identify a variety of regularisation methods, among them classical norm penalties (L1, L2, Elastic Net), stochastic regularisation methods (dropout and its iterations), Bayesian regularisation, feature selection methods, and optimization-related regularizers, and domain-specific hybrid methods. The study methods across the surveyed literature have proved to be able to benefit generalisation performance, increase robustness to noisy and limited data, stabilise the training process, and in some instances, benefit model interpretability.

It is revealed that there is no single best regularisation algorithm to use; rather, the efficiency of either methodology varies depending on the nature of the data, the model structure, and the field of application. As an example, L1 penalties are useful in high-dimensional feature selection problems, dropout versions are useful in deep vision models, and the use of Bayesian regularisation is especially useful in the field of engineering and inverse problems. Many other studies also highlighted the interaction between explicit and implicit regularisation and indicated as well that training algorithms like stochastic gradient descent naturally influence generalisation behaviour. The evidence that has been reviewed indicates that a combination of complementary data-level, model-level and algorithmic regularizers will lead to a performance that is most optimal especially in a real-life environment where noise, complexity and scarcity of data exist simultaneously.

To sum it up, regularisation is an unavoidable approach to the development of predictive systems by which one can be sure that the system is reliable in the age of more complicated models. It is possible that future principled development of models can be informed by a better understanding of the interaction between various regularization techniques with model capacity, optimization dynamics, and dataset characteristics. This work would serve that purpose by bringing together the existing theoretical knowledge and empirical evidence that would provide practitioners and researchers with a base on which to use when making informed decisions concerning regularisation strategies. Future studies ought to be based on comparative assessment of a variety of tasks, automated regularisation selection, and combination of adaptive and data-driven regularizers, which are able to strike a balance between model expressivity and generalisation in dynamically changing environments.

6. Future Directions

6.1. Understanding Interactions Between Explicit and Implicit Regularization

Although explicit methods like L1/L2 penalties and dropout are already adequately investigated, the implicit regularization processes that occur because of optimization algorithms (e.g. SGD, momentum, adaptive optimizers) are still not sufficiently understood. To better understand the interaction between these two types of regularisations, it will be beneficial to develop more theoretical sociological models to understand why overparameterized models can provide effective generalisation.

6.2. Comparative Evaluations Across Diverse Data Regimes

Existing empirical results tend to be domain-dependent thus creating gaps in the learning behaviour of regularisation methods in extreme situations, e.g. when noise is high, when samples are limited, or when there is a strong class imbalance. Overall systematic benchmarks that consider regularizers on various data complexities and modalities would offer more broadly applicable advice to practitioners.

6.3. Development of Adaptive and Context-Aware Regularization Methods

Another promising direction is adaptive regularization schemes, in which model constraints are dynamically changed according to training signals, sample uncertainty, or changing data distributions. Developing adaptive algorithms which adaptively estimate the strength of regularisation may provide a major boost in the generalisation capabilities in non-stationary settings.

6.4. Hybrid Regularization Frameworks for Complex Applications

Statistical, Bayesian, and deep learning-based regularisation techniques may be combined to provide a promising direction of creating better and more interpretable systems. Hybridized frameworks combining feature selection, parameter penalties, dropout forms and probabilistic priors could be of great value to high-stakes fields like healthcare, cybersecurity, remote sensing, and the design of engineering.

6.5. New Diagnostic Tools for Non-Classical Generalization Phenomena

The onset of double-descent risk curves as well as other atypical generalisation patterns critiques the traditional practise of diagnosis. It is increasingly being real-

ised that better tools, measures and visualisation are required to help researchers identify, interpret and reduce overfitting in such complicated regimes. This type of diagnostics would be used in the more effective selection and fine-tuning of regularisation strategies.

6.6. Automated Regularization Selection and Optimization

With the increase in complexity of model architectures, manual tuning of regularisation hyperparameters is no longer feasible. The automated methods, including AutoML-assisted regularizer choice, neural architecture search with intrinsic regularization, and the reinforcement learning-based actions suggest an opportunity to reduce human involvement and maintain a high level of generalization.

7. Declarations

7.1. Author Contributions

Osita Miracle Nwakeze: Writing - Original Draft, Writing - Review & Editing, Conceptualization, Resources.

7.2. Institutional Review Board Statement

Not applicable.

7.3. Informed Consent Statement

Not applicable.

7.4. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

7.5. Acknowledgment

Not applicable.

7.6. Conflicts of Interest

The authors declare no conflicts of interest.

8. References

- [1] C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals, "Understanding deep learning requires rethinking generalization," *arXiv preprint arXiv:1611.03530*, 2016. [Online]. Available: <https://arxiv.org/abs/1611.03530>
- [2] Y. Tian and Y. Zhang, "A comprehensive survey on regularization strategies in machine learning," *Information Fusion*, vol. 80, pp. 146–166, 2022, <https://doi.org/10.1016/j.inffus.2021.11.005>.
- [3] I. S. C. Umerah, U. A. Tochukwu, A. Maduabuchuku, O. M. Nwakeze, and O. C. Akachukwu, I. S. Chinasa "AI-driven diagnostic imaging: Hybrid CNN-GNN models for early detection of cancer from pathological images," *Int. J. Latest Technol. Eng., Manag. Appl. Sci.*, vol. 14, no. 9, pp. 579–588, 2025. <https://doi.org/10.51583/IJLTEMAS.2025.1409000070>.
- [4] R. Moradi, R. Berangi, and B. Minaei-Bidgoli, "A survey of regularization strategies for deep models," *Artif. Intell. Rev.*, vol. 53, no. 6, pp. 3947–3986, 2020, <https://doi.org/10.1007/s10462-019-09784-7>.

- [5] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, pp. 1929–1958, 2014. <https://dl.acm.org/doi/abs/10.5555/2627435.2670313>.
- [6] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, San Francisco, CA, USA, 2016, pp. 785–794. <http://dx.doi.org/10.1145/2939672.2939785>.
- [7] J. Kolluri, V. K. Kotte, M. S. B. Phridviraj, and S. Razia, "Reducing overfitting problem in machine learning using novel L1/4 regularization method," in *Proc. 4th Int. Conf. Trends Electron. Informatics (ICOEI)*, IEEE, 2020. <https://doi.org/10.1109/ICOEI48184.2020.9143022>.
- [8] A. Pak, A. K. Rad, M. J. Nematollahi, and M. Mahmoudi, "Application of the Lasso regularisation technique in mitigating overfitting in air quality prediction models," *Sci. Rep.*, vol. 15, art. no. 547, 2025. <https://doi.org/10.1038/s41598-024-84342-y>.
- [9] G. C. Cawley and N. L. C. Talbot, "Preventing over-fitting during model selection via Bayesian regularisation of the hyper-parameters," *J. Mach. Learn. Res.*, vol. 8, pp. 841–861, 2007. <https://dl.acm.org/doi/abs/10.5555/1248659.1248690>.
- [10] M. Ghayoumi, "Enhancing efficiency and regularization in convolutional neural networks: Strategies for optimized dropout," *AI*, vol. 6, art. no. 111, 2025. <https://doi.org/10.3390/ai6060111>.
- [11] X. Ying, "An overview of overfitting and its solutions," *J. Phys.: Conf. Ser.*, vol. 1168, no. 2, art. no. 022022, 2019. <https://doi.org/10.1088/1742-6596/1168/2/022022>.
- [12] P. Zhang, S. Zhang, X. Liu, L. Qiu, and G. Yi, "A least squares ensemble model based on regularization and augmentation strategy," *Appl. Sci.*, vol. 9, art. no. 1845, 2019. <https://doi.org/10.3390/app9091845>.
- [13] I. Marin, A. Kuzmanic Skelin, and T. Grujic, "Empirical evaluation of the effect of optimization and regularization techniques on the generalization performance of deep convolutional neural network," *Appl. Sci.*, vol. 10, art. no. 7817, 2020. <https://doi.org/10.3390/app10217817>.
- [14] A. Kanavos, M. Trigka, E. Dritsas, G. Vonitsanos, and P. Mylonas, "A regularization-based big data framework for winter precipitation forecasting on streaming data," *Electronics*, vol. 10, art. no. 1872, 2021. <https://doi.org/10.3390/electronics10161872>.
- [15] T. Kotsilieris, I. Anagnostopoulos, and I. E. Livieris, "Regularization techniques for machine learning and their applications," *Electronics*, vol. 11, art. no. 521, 2022. <https://doi.org/10.3390/electronics11040521>.
- [16] A. B. Shanmugavel *et al.*, "A novel ensemble-based reduced overfitting model with convolutional neural network for traffic sign recognition system," *Electronics*, vol. 12, art. no. 926, 2023. <https://doi.org/10.3390/electronics12040926>.
- [17] F. Jin, X. Chen, Y. Yu, and K. Li, "An improved regularization stochastic configuration network for robust wind speed prediction," *Energies*, vol. 18, art. no. 6170, 2025. <https://doi.org/10.3390/en18236170>.
- [18] A. Mohammad-Djafari, "Regularization, Bayesian inference, and machine learning methods for inverse problems," *Entropy*, vol. 23, art. no. 1673, 2021. <https://doi.org/10.3390/e23121673>.
- [19] S. Gupta, R. Gupta, M. Ojha, and K. P. Singh, "A comparative analysis of various regularization techniques to solve overfitting problem in artificial neural network," in *Proc. REDSET 2017*, CCIS, vol. 799, Springer, 2018, pp. 363–371. https://doi.org/10.1007/978-981-10-8527-7_30.
- [20] F. Ahmadi Abkenari, A. Milani Fard, and S. Khanchi, "Hybrid machine learning-based approaches for feature and overfitting reduction to model intrusion patterns," *J. Cybersecurity Privacy*, vol. 3, no. 3, pp. 544–557, 2023. <https://doi.org/10.3390/jcp3030026>.
- [21] X. Zhang, L. Sun, and L. Qi, "Bayesian regularization algorithm based recurrent neural network method and NSGA-II for the optimal design of the reflector," *Machines*, vol. 10, art. no. 63, 2022. <https://doi.org/10.3390/machines10010063>.
- [22] M. Liu, G. Srivastava, J. Ramanujam, and M. Brylinski, "Insights from augmented data integration and strong regularization in drug synergy prediction with SynerGNet," *Mach. Learn. Knowl. Extr.*, vol. 6, no. 3, pp. 1782–1797, 2024. <https://doi.org/10.3390/make6030087>.
- [23] M. Xiao *et al.*, "Addressing overfitting problem in deep learning-based solutions for next generation data-driven networks," *Wireless Commun. Mobile Comput.*, vol. 2021, art. ID 8493795, pp. 1–10, 2021. <https://doi.org/10.1155/2021/8493795>.
- [24] C. F. G. Santos and J. P. Papa, "Avoiding overfitting: A survey on regularization methods for convolutional neural networks," *ACM Trans. Comput. Syst.*, vol. 1, no. 1, pp. 1–27, 2018. <https://doi.org/10.1145/1122445.1122456>.
- [25] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006. <https://link.springer.com/9780387310732>.

- [26] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016. <https://synapse.koreamed.org/pdf/10.4258/hir.2016.22.4.351>.
- [27] M. Belkin, D. Hsu, S. Ma, and S. Mandal, "Reconciling modern machine-learning practice and the classical bias–variance trade-off," *Proc. Natl. Acad. Sci. USA*, vol. 116, no. 32, pp. 15849–15854, 2019, <https://doi.org/10.1073/pnas.1903070116>.