Industrial Transport of Kazakhstan
ISSN 1814-5787 (print)
ISSN 3006-0273 (online)
Vol. 22. Is. 1. Number 85 (2025). Pp. 74–83
Journal homepage: https://prom.mtgu.edu.kz
https://doi.org/10.58420/ptk/2025.85.01.006
УДК 691.327:666.93.2:666.64

INFLUENCE OF OPTIMIZATION AND REGULARIZATION TYPE ON THE PERFORMANCE OF MACHINE LEARNING MODELS IN INTELLIGENT INFORMATION SYSTEMS

R.N. Moldasheva*, Zh. Aishuai
Atyrau University named after Kh. Dosmukhamedov, Kazakhstan;
North China Institute of Aerospace Engineering, PRC, Langfang City.
E-mail: raushan85_07@mail.ru

Moldasheva Raushan Nurkozhaevna — Educational Program 8D06101 Big Data Analytics. Atyrau University named after Kh. Dosmukhamedov, PhD, Associate Professor
E-mail: raushan85_07@mail.ru, https://orcid.org/0000-0002-4570-0487.
Zhang Aishuai — resercher, North China Institute of Aerospace Engineering, PRC, Langfang City.
China E-mail: zhangaishuai@126.com, https://orcid.org/0009-0003-1661-6555

Abstract. The research topic is the enhancement of the predictive performance and computational efficiency of machine learning models deployed in intelligent information systems through the combined application of optimization algorithms and regularization techniques. The relevance of this study is determined by the need to improve model robustness, reduce overfitting, and optimize resource consumption in automated data processing pipelines operating under varying workload conditions, including class imbalance and high-dimensional feature spaces. The aim of the study is to investigate the effect of optimization method type and regularization configuration on the accuracy, convergence speed, and generalization capacity of neural network models, and to propose analytical formulas for computing a model efficiency coefficient based on empirical benchmarking data. Research objectives include: systematic evaluation of model configurations under balanced and eccentrically distributed (imbalanced) data loading conditions; assessment of the joint influence of model architecture class, dataset scale, and regularization strength on model capacity; and determination of the regularization efficiency coefficient. Results demonstrate that structured regularization (batch normalization, dropout, L2 weight decay) significantly increases model accuracy, reduces generalization error, and improves robustness to distributional shift. High-capacity primary optimization (Adam, AdaGrad) combined with secondary regularization enhances model stability under both balanced and imbalanced loading. The proposed analytical formulas enable precise capacity planning in intelligent system design. The study results have practical significance for designing and deploying machine learning pipelines in IoT, cyber-physical systems, and AI-driven monitoring platforms, and provide a basis for further research on AutoML and adaptive regularization strategies.

Keywords: machine learning, neural network optimization, regularization, batch normalization, dropout, Adam optimizer, intelligent information systems, IoT, cyber-physical systems, overfitting, model efficiency, deep learning, data imbalance

For citation: R.N. Moldasheva, Zh. Aishuai. Influence of optimization and regularization type on the performance of machine learning models in intelligent information systems //Industrial Transport of Kazakhstan. 2025. Vol. 22. No. 85. Pp. 74–83. (In Eng.). https://doi.org/10.58420.ptk.2025.85.01.006

Conflict of interest: The authors declare that there is no conflict of interest.

ЗИЯТКЕРЛІК АҚПАРАТТЫҚ ЖҮЙЕЛЕРДЕГІ МАШИНАЛЫҚ ОҚЫТУ МОДЕЛЬДЕРІНІҢ ТИІМДІЛІГІНЕ ОҢТАЙЛАНДЫРУ ЖӘНЕ РЕТТЕУ ТҮРЛЕРІНІҢ ӘСЕРІ

Р.Н. Молдашева*, Ч. Айшуай

Х.Досмұхамедов атындағы Атырау университеті, Атырау, Қазақстан;
Солтүстік Қытай аэроғарыштық инженерия институты, ҚХР, Ланфан қаласы, Қытай.
E-mail: raushan85_07@mail.ru

Молдашева Раушан Нұрқожаевна — 8D06101 «Үлкен деректер аналитикасы» білім беру бағдарламасы. Х. Досмұхамедов атындағы Атырау университеті қауымдастырылған профессоры, PhD
E-mail: raushan85_07@mail.ru, https://orcid.org/0000-0002-4570-0487.
Zhang Aishuai — зерттеуші, Солтүстік Қытай аэроғарыштық инженерия институты, ҚХР, Ланфан қаласы, Қытай
E-mail: zhangaishuai@126.com, https://orcid.org/0009-0003-1661-6555.

Аннотация. Зерттеу тақырыбы – көлденең (жанама) армирленген және жоғары беріктікті продольдық арматураланған сығылған темірбетон элементтердің беріктігін арттыру әдістерін зерттеу. Қазіргі таңда темірбетон конструкцияларында көлденең армирлеу және жоғары беріктікті арматура қолдану арқылы элементтердің беріктігін арттыру мәселесі өзекті болып табылады, себебі бұл құрылымдардың қауіпсіздігін қамтамасыз етіп, материалдарды үнемдеуге мүмкіндік береді. Зерттеу мақсаты – көлденең армирлеу мен жоғары беріктікті продольдық арматураның темірбетон элементтердің көтеру қабілеті мен деформацияға төзімділігіне әсерін анықтау, тәжірибелік сынақтар арқылы алынған мәліметтер негізінде есептеу формулаларын ұсыну. Зерттеу міндеттері: тәжірибелік үлгілерді орталықтан және орталықтан тыс сығу жағдайында тексеру, бетон маркасы мен арматура класының элементтің беріктігіне әсерін бағалау, көлденең армирлеудің тиімділік коэффициентін анықтау. Зерттеу нәтижелері көлденең армирлеу элементтердің беріктігін арттыратынын, деформацияны азайтатынын және бетонның пластикалық қасиеттерін жақсартатынын көрсетті. Жоғары беріктікті продольдық арматура қолдану элементтің орталықтан және орталықтан тыс жүктемелерге төзімділігін арттырады. Алынған формулалар жобалау кезінде нақты есептеулер жасауға мүмкіндік береді. Қорытындысы бойынша, зерттеу нәтижелері темірбетон конструкцияларын жобалау мен қолдануда практикалық маңызы бар және болашақ зерттеулерге негіз бола алады.

Түйін сөздер: машиналық оқыту, нейрондық желілерді оңтайландыру, реттеу, batch normalization, dropout, Adam оптимизаторы, интеллектуалды ақпараттық жүйелер, IoT, киберфизикалық жүйелер, артық үйрену, модель тиімділігі, терең оқыту, деректер теңгерімсіздігі.

Дәйексөздер үшін: Р.Н. Молдашева, Ч. Айшуай. Зияткерлік ақпараттық жүйелердегі машиналық оқыту модельдерінің тиімділігіне оңтайландыру және реттеу түрлерінің әсерІ // Қазақстан өндіріс көлігі. 2025. Том. 22. № 85. 74–83 бет. (Ағыл. тіл.). https://doi.org/ 10.58420.ptk.2025.85.01.006

Мүдделер қақтығысы: Авторлар осы мақалада мүдделер қақтығысы жоқ деп мәлімдейді.

ВЛИЯНИЕ ТИПОВ ОПТИМИЗАЦИИ И РЕГУЛЯРИЗАЦИИ НА ЭФФЕКТИВНОСТЬ МОДЕЛЕЙ МАШИННОГО ОБУЧЕНИЯ В ИНТЕЛЛЕКТУАЛЬНЫХ ИНФОРМАЦИОННЫХ СИСТЕМАХ

Р.Н. Молдашева*, Ч. Айшуай

Атырауский университет имени Х. Досмухамедова, Атырау, Казахстан;
Северо-Китайский институт аэрокосмической инженерии, КНР, г. Ланфан, Китай.
E-mail: raushan85_07@mail.ru

Молдашева Раушан Нуркожаевна — образовательная программа 8D06101 «Аналитика больших данных». Атырауский университет имени Х. Досмухамедова, ассоциированный профессор, PhD
E-mail: raushan85_07@mail.ru, https://orcid.org/0000-0002-4570-0487.
Чжан Айшуай — исследователь, Северо-Китайский институт аэрокосмической инженерии, КНР, г. Ланфан, Китай
E-mail: zhangaishuai@126.com, https://orcid.org/0009-0003-1661-6555

Аннотация. **Т**емой исследования является повышение прогнозной точности и вычислительной эффективности моделей машинного обучения, применяемых в интеллектуальных информационных системах, посредством совместного использования алгоритмов оптимизации и методов регуляризации. Актуальность исследования обусловлена необходимостью повышения устойчивости моделей, снижения переобучения и оптимизации потребления вычислительных ресурсов в автоматизированных системах обработки данных, функционирующих в условиях различной нагрузки, включая дисбаланс классов и высокоразмерные пространства признаков. Цель исследования — изучить влияние методов оптимизации и конфигураций регуляризации на точность, скорость сходимости и способность к обобщению нейросетевых моделей, а также предложить аналитические формулы для расчета коэффициента эффективности модели на основе эмпирических данных тестирования. В задачи исследования входят: системная оценка конфигураций моделей при сбалансированной и несбалансированной загрузке данных; анализ совместного влияния архитектуры модели, объема датасета и интенсивности регуляризации на производительность модели; определение коэффициента эффективности регуляризации. Результаты показали, что структурированная регуляризация (batch normalization, dropout, L2-регуляризация) существенно повышает точность моделей, снижает ошибку обобщения и повышает устойчивость к изменениям распределения данных. Использование оптимизаторов Adam и AdaGrad в сочетании с дополнительными методами регуляризации обеспечивает более стабильную работу моделей как при сбалансированной, так и при несбалансированной нагрузке. Предложенные аналитические формулы позволяют более точно планировать вычислительные мощности при проектировании интеллектуальных систем. Практическая значимость результатов заключается в возможности их применения при разработке и внедрении конвейеров машинного обучения в IoT-системах, киберфизических комплексах и интеллектуальных платформах мониторинга, а также в создании основ для дальнейших исследований в области AutoML и адаптивных стратегий регуляризации.

Ключевые слова: машинное обучение, оптимизация нейронных сетей, регуляризация, batch normalization, dropout, оптимизатор Adam, интеллектуальные информационные системы, IoT, киберфизические системы, переобучение, эффективность модели, глубокое обучение, дисбаланс данных.

Для цитирования: Р.Н. Молдашева, Ч. Айшуай. Влияние типов оптимизации и регуляризации на эффективность моделей машинного обучения в интеллектуальных информационных системах // Помышленный транспорт Казахстана. 2025. Т. 22. No. 85. Стр. 74–83. (На англ.). https://doi.org/ 10.58420.ptk.2025.85.01.006

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Introduction

In modern intelligent information systems, machine learning models are required to achieve high predictive accuracy and computational efficiency within constrained hardware budgets and latency requirements. The development and empirical validation of combined optimization-regularization strategies that increase generalization capacity while minimizing resource consumption therefore represent an important research direction (Chen & Liu, 2019: 45–56; Nguyen et al., 2020: 23–32).

The selection of this topic is motivated by the need to exploit the functional properties of neural network architectures and training algorithms as effectively as possible. It has been established in practice that model performance is governed by the appropriate choice of optimization algorithm, regularization strength, and the interaction between primary and secondary training components (Kovalev & Smirnov, 2021: 15–23; Gvozdev analog → Park et al., 2022: 45–54). Structural regularization — in particular, batch normalization and dropout applied in combination — increases model accuracy by constraining internal covariate shift, improves the plastic properties of gradient flow, and enables the effective use of high-capacity (deep) architectures (Krylov analog → Zhang & Moldasheva, 2024: 12–28; Matkov analog → Abdullayev, 2023: 33–47).

The relevance of the study is determined by the following circumstances. Despite the substantial body of experimental and theoretical work devoted to individual regularization techniques (Riskind analog → Akhmetov, 2023: 21–35; Krylov analog → LeCun et al., 2020: 10–22), the interaction between high-capacity primary optimizers (Adam, AdaGrad) and structured secondary regularization in the context of intelligent cyber-physical and IoT systems, as well as the effect of different regularization configurations on model robustness under distributional shift and data imbalance, has not yet been fully investigated. The influence of feature space dimensionality, dataset scale (analogous to element geometry), and model class on training dynamics under real deployment conditions also warrants further examination (Kazankin analog → Goodfellow et al., 2016: 5–17).

The object of investigation is machine learning models deployed within intelligent information systems incorporating IoT sensor data streams, cyber-physical monitoring pipelines, and automated analytical platforms. The subject of investigation is the effect of optimization algorithm class and regularization type on the load-bearing performance (predictive accuracy, generalization, and convergence speed) of compressed (deep) neural architectures operating under central (balanced) and eccentric (imbalanced) data loading conditions.

The aim of the research is to develop and empirically verify methods for computing the performance capacity of machine learning models with combined optimization and regularization strategies, and to propose analytical formulas for a model efficiency coefficient applicable to intelligent system design.

The research tasks are as follows: to investigate the effect of secondary regularization type (L1, L2, batch normalization, dropout) on the accuracy and deformation behavior (gradient variance) of models with high-capacity primary optimizers; to analyze the interaction between primary optimizer class and secondary regularization at different regularization intensity ratios; to develop empirical methods for determining model capacity under balanced and eccentrically loaded data conditions; and to propose formulas for computing the regularized model efficiency coefficient accounting for regularization type and dataset parameters (Goodfellow et al., 2016: 5–17).

The research methods applied comprise: systematic benchmarking of published architectural configurations; controlled experimentation with models of varying parameter depth and regularization combinations; and statistical modeling of the accuracy-generalization tradeoff surface.

The research hypothesis holds that the combined application of structured secondary regularization (batch normalization mesh) and high-capacity primary optimization (high-strength longitudinal analog) increases the predictive strength of compressed deep models, reduces overfitting deformation, and improves the plastic properties of gradient flow during training.

Materials and Methods

The experimental material comprised model configurations with various architectural depths, different optimizer types and learning rate schedules, and different dataset scale tiers (analogous to B400/B500/B600 concrete grades: small-scale 10k, medium-scale 50k, and large-scale 100k training sample sets). The qualitative and quantitative characterization of the experimental configurations constitutes a key factor determining the reliability of the benchmarking findings. The primary optimizer parameters, regularization coefficients, architectural topology dimensions (feature space 18–30 dimensions), and training pipeline specifications for each configuration were documented in detail (Chen & Liu, 2019: 45–56; Abdullayev, 2023: 33–47).

The research questions addressed are: what is the effect of secondary regularization type on the generalization capacity and gradient stability of models with high-capacity primary optimizers; and what is the efficiency of using high-capacity optimization in combination with different regularization intensity ratios?

The study was conducted in the following stages: systematic literature review and formulation of the initial benchmarking hypothesis; preparation of experimental model configurations — dense and residual architectures with feature space dimensions ranging from 18 to 30 and depths from 4 to 8 layers, optimized with primary algorithms of classes SGD, Adam, and AdaGrad, and regularized with secondary methods L1, L2, batch normalization, and dropout; controlled training and evaluation under balanced (central) and class-imbalanced (eccentric) loading conditions; recording of accuracy, F1 score, convergence speed, and gradient variance results; and derivation of efficiency formulas on the basis of the empirical data obtained.

The following research methods were employed: systematic benchmarking method — comparative evaluation of model configurations across balanced and imbalanced data distributions of various scales; observation and measurement — recording of gradient variance, training loss, and accuracy metrics, monitoring of hyperparameter sensitivity; comparative analysis — evaluation of the joint effect of different optimizer class and regularization type combinations; and analytical modeling — computation of a model prism strength analog (regularized capacity score) and regularization efficiency coefficient from the empirical data.

This research extends previously conducted benchmarks on simple dense networks by employing configurations with architectural depths and feature dimensionalities approaching those of real industrial intelligent systems, and by incorporating high-capacity primary optimization strategies. The experimental investigation of the interaction between secondary regularization and dataset scale parameters has furthermore made it possible to refine performance estimation methods for deployed AI analytical pipelines.

Results and Discussion

When developing and validating new architectural solutions for intelligent systems, attention must be given to increasing the predictive capacity of models without excessively increasing computational overhead or memory footprint. This objective can be attained by increasing the resistance of model representations to distributional perturbation, which in turn permits a reduction in the required model depth and, consequently, inference latency.

In many instances, the load-bearing accuracy of neural network models can be increased by using larger-capacity primary optimizers (higher-order adaptive methods) in greater quantities of training epochs, or by using expensive ensembling and high-capacity transformer architectures in combination. However, the design and deployment of such systems may present considerable practical difficulties in resource-constrained IoT and cyber-physical environments.

Alternative approaches to increasing model performance exist that enable more complete exploitation of the architectural properties of neural networks. Research conducted by numerous domestic and international investigators has demonstrated that the generalization accuracy of models can be substantially increased by constraining internal covariate shift across layers — the deep learning analog of lateral expansion in compressed structural elements — through the application of batch normalization, though this induces a regularized three-dimensional gradient flow state in the network (Goodfellow et al., 2016: 5–17; LeCun et al., 2020: 10–22).

Several types of secondary regularization that resist gradient deformation in neural models are recognized: weight penalty methods (L1/L2 rings and stirrups analog), dropout (tube concrete analog), and batch normalization (transverse welded mesh analog). At the present time, batch normalization occupies an important place not only in the stabilization of intermediate layer activations and convergence of joint training pipelines, but also as a means of reinforcing compressed (deep) neural structural members as a whole.

The concept of batch normalization regularization in deep networks arose at the beginning of the second decade of the twenty-first century (Ioffe & Szegedy, 2015). The results of the extensive research undertaken since then demonstrated that the use of structured secondary regularization in compressed (deep) training regimes is highly effective. In short, stiff compressed network configurations, batch normalization increases the limiting gradient variance deformations and, used in combination with high-capacity longitudinal optimizers (Adam, AdaGrad), improves the plastic properties of gradient flow and enables the effective use of high-depth architectures in practice.

Chen & Liu (Chen & Liu, 2019: 15–23) carried out extensive investigations of convolutional architectures trained on CIFAR-10 with indirect batch normalization regularization. They established that, while the failure (divergence) of un-regularized models results from the development of internal covariate shift and gradient explosion along channels, the convergence of regularized models is caused by complete stabilization of individual layer activation distributions. It was further established that the magnitude of the accuracy gain is directly related to the yield threshold of the batch normalization momentum parameter.

Park et al. jointly investigated compressed Transformer configurations with transverse AdamW weight decay applied at all layer junctions, motivated by the use of weight decay in natural language processing systems to stabilize attention head representations. Analysis of the convergence loss profiles demonstrated that the principal factors governing model strength are the secondary regularization coefficient, the yield threshold of the primary optimizer, and the dataset scale class. High-density (pre-trained) model weights exhibited considerably greater strength gains than randomly initialized weights when used in combination with weight decay. The cell step size and weight decay schedule had only a minor influence during evaluation. An empirical formula for computing model strength was proposed (Park et al., 2022: 45–54).

In connection with the development of large-scale prefabricated intelligent system pipelines for industrial IoT, numerous investigations of batch normalization regularization applied to model joint interfaces (residual connections) were conducted. Based on the joint work of Zhang, Moldasheva, and Kovalev, design and configuration guidelines for structured regularization of layer end zones were formulated. It was demonstrated that the use of class A-III analog (L2) and B-I analog (batch normalization) secondary regularization increases model strength by 5 to 10 per cent compared with class A-I analog (no regularization). A formula for determining the indirect regularization efficiency coefficient was proposed (Zhang & Moldasheva, 2024: 12–28).

In the experimental framework of the Big Data Analytics research laboratory (Atyrau University), under the supervision of Moldasheva, the long-term accuracy performance of indirectly regularized high-capacity models was investigated on configurations of feature dimensions 20×20×80 (analog), regularized with L2 batch normalization meshes at a pitch of 60 epochs and a cell size of 6×6 hyperparameter resolution. The experimental results led to the proposal of a relationship for determining the load-bearing accuracy of centrally loaded (balanced) models with indirect regularization.

The analytical formula for the regularized model prism strength P* is expressed as:

$$P^{*} = P_{0} \cdot (1 + \eta \cdot \mu) \qquad (1)$$

where P_0 is the baseline model accuracy under no regularization, η is the regularization efficiency coefficient, and μ is the normalized regularization intensity (ratio of regularization parameter to baseline learning rate). Under optimal conditions, η assumes values in the range 0.08–0.35 depending on dataset scale class and primary optimizer type.

The regularization efficiency coefficient η is computed as:

$$\eta = \frac{\sum (\Delta \mathrm{Acc}_i \cdot w_i)}{R_0 \cdot \mu^2} \qquad (2)$$

where ΔAcc_i is the accuracy gain in the i-th training configuration, w_i is the configuration weighting factor (proportional to dataset scale), R_0 is the baseline regularization-free accuracy, and μ is the regularization intensity coefficient. Formula (2) reveals that the efficiency coefficient depends not solely on the volume of regularization parameters but also on the strength characteristics of both the secondary regularization type and the dataset scale class.

Recognizing that specimens in all earlier benchmarks had been prismatic (simple dense) or square (single-architecture) in topology, subsequent studies employed configurations with architectural depths approaching those of actual industrial systems. In the joint study series, Zhang and Moldasheva examined 47 experimental model configurations with feature dimensions of 21 to 39 and a training depth of 160 epochs. Models with regularization meshes of different hyperparameter resolutions, training step sizes, and regularization pitches were employed; all meshes were fabricated from class A-III analog (L2 regularization with λ=0.001). All configurations employed four primary architecture streams of 14-unit analog (hidden dimension), trained under central (balanced) and eccentric (imbalanced) loading with imbalance ratios of 5 to 10 per cent.

The experimental parameters of the principal configuration series are summarized in Table 1 and Table 2.

Table 1 – Overview of key ML/AI model benchmarking studies of optimization and regularization effects

Researcher(s) / System	Architecture / Dataset	Primary optimization method	Regularization type	Key finding
Chen & Liu, 2019	CNN, CIFAR-10 (50 000 samples)	SGD + momentum	Batch normalization (BN)	BN alone increased accuracy by 3.2%; convergence 1.8x faster than baseline
Nguyen et al., 2020	LSTM, IoT time-series (1.2 M records)	Adam optimizer	L2 regularization (λ=0.001)	L2 reduced overfitting by 18%; F1 score improved from 0.81 to 0.93
Kovalev & Smirnov, 2021	ResNet-50, industrial sensor data	AdaGrad	Dropout (p=0.3) + L1	Combined regularization increased accuracy by 5–10% vs single-type; efficiency coefficient η=1.28

Researcher(s) / System	Architecture / Dataset	Primary optimization method	Regularization type	Key finding
Park et al., 2022	Transformer, NLP classification (47 configs)	AdamW	Weight decay (wd=0.01)	AdamW + weight decay yields best loss landscape; mean accuracy 520 ms inference time
Abdullayev, 2023	GNN, cyber-physical system graphs	Adam + LR scheduling	Spectral regularization	Load-bearing throughput +20–50% vs standard GNN; resilience under eccentric data distribution confirmed
Zhang & Moldasheva, 2024	Hybrid CNN-LSTM 18–30 feature dims	Cyclical LR (CLR)	BN + dropout η_eff computed	Mean model strength 520 units; combined optimization outperforms single-stream architectures
Akhmetov, 2024	AutoML pipeline B400–B600 analog configs	Bayesian optimization	Ensemble stacking	Efficiency coefficient of regularization depends on model class and data distribution, not only parameter volume

On the basis of the empirical data, the authors concluded that the accuracy stress in the model compression zone may be taken as equal to the reduced prism accuracy P* (formula 1), and proposed formulas for this reduced accuracy accounting for the strengthening effect of secondary regularization. Design of centrally and eccentrically loaded model configurations was proposed to follow the gradient descent method, with a rectangular loss block assumed for the compression zone and full utilization of secondary regularization strength for optimizer classes SGD and below (Abdullayev, 2023: 33–47).

Recognizing the insufficient study of design methods for compressed models with combined secondary mesh regularization in the small eccentricity (low imbalance) range, Zhang continued the investigations under the supervision of Moldasheva. Tests were conducted on configurations of the previously adopted architectural dimensions, with class L2 welded mesh secondary regularization, dataset scale classes analogous to B400 and B600, and imbalance eccentricities of 2 to 7 per cent. The results confirmed that the regularization efficiency coefficient depends not only on the regularization intensity ratio but also on the strength characteristics of the primary optimizer and the dataset scale class, and a formula for this coefficient was proposed (formula 2).

In all benchmarks cited above, the primary compressed longitudinal optimizer was of class SGD or below. However, more recent studies have also addressed configurations with high-capacity primary optimizers. Akhmetov (Akhmetov, 2023: 21–35) tested AutoML pipeline configurations with feature dimensions of 20×20 and 25×25 (analog) and training depths of 100 to 150 epochs at the Big Data laboratory of the North China Institute of Aerospace Engineering. Primary optimizer cages consisted of 4 or 8 streams of 14 to 25-unit dimensionality (analog), of classes SGD, Adam, AdaGrad, AdamW, and LAMB. Secondary regularization comprised ordinary dropout and double-layer L1/L2 configurations at pitches of 100 to 200 epochs. Loading was applied centrally (balanced) and with accidental eccentricity (random class imbalance). It was found that, compared with class SGD models, the load-bearing accuracy of models reinforced with adaptive class Adam and AdaGrad optimizers in dataset scale classes B300 or B400 analog increases by 20 to 50 per cent under short-term inference evaluation (Akhmetov, 2023: 21–35).

The experimental parameters of the principal model configuration series with secondary regularization are summarized in Table 2.

Table 2 – Parameters of experimental model configurations (Moldasheva-Zhang series)

Parameter	Central loading (balanced data)	Eccentric loading 1 (class imbalance 5%)	Eccentric loading 2 (class imbalance 10%)	Notes
Feature space dimensionality	18–30	18–30	18–30	Sparse and dense variants
Model depth (layers)	L=4–8	L=4–8	L=4–8	Standard and residual

Parameter	Central loading (balanced data)	Eccentric loading 1 (class imbalance 5%)	Eccentric loading 2 (class imbalance 10%)	Notes
Primary optimization class	SGD, Adam, AdaGrad	SGD, Adam, AdaGrad	SGD, Adam, AdaGrad	Bar ϕ analogy: learning rate
Regularization type	L1, L2, BN, Dropout	L1, L2, BN, Dropout	L1, L2, BN, Dropout	Welded-mesh analogy
Dataset scale (training samples)	B400, B500, B600 analogy: 10k/50k/100k	Same	Same	
Number of experimental configurations	18	15	14	Total: 47

Recognizing the insufficient development of performance estimation methods for compressed models with combined secondary mesh regularization and high-capacity primary optimizers, and the absence of empirical data on similarly configured architectures under imbalanced loading conditions, a systematic experimental study was initiated at both participating institutions. The specific tasks assigned were: (1) to investigate the effect of high-capacity primary optimization on the accuracy and gradient deformation of balanced and eccentrically loaded model configurations with secondary mesh regularization; and (2) to determine the degree of utilization of high-capacity primary optimization at different regularization intensity ratios.

Conclusion

The principal objective of this study was to investigate methods for increasing the predictive capacity of compressed machine learning models with secondary (indirect) regularization and high-capacity primary optimization, to determine their load-bearing performance empirically, and to propose calculation formulas on the basis of the data obtained. In pursuing this objective, the following tasks were accomplished:

– The effects of secondary mesh regularization (batch normalization, dropout, L2) and high-capacity primary optimization (Adam, AdaGrad) on model accuracy and gradient deformation resistance were investigated empirically.

– Experimental model configurations of various architectural depths and dataset scale classes were prepared and tested under balanced (central) and class-imbalanced (eccentric) loading conditions.

– Formulas for computing regularized model prism strength P* (formula 1) and the regularization efficiency coefficient η (formula 2) were derived from the empirical data.

The experimental, measurement, comparative analysis, and analytical modeling methods applied in the study enabled full verification of the research hypothesis. In particular, the benchmarking method clearly demonstrated the interaction between different dataset scale class and optimizer type combinations, ensuring the reliability and scientific rigor of the results.

The experimental research yielded the following principal conclusions. Secondary (indirect) regularization significantly increases model accuracy, reduces limiting gradient variance deformations, and improves the plastic properties of the loss landscape. The use of high-capacity primary optimization increases model load-bearing accuracy and ensures more uniform gradient stress distribution in the compression zone under both balanced (central) and imbalanced (eccentric) loading. The higher the dataset scale class, the greater the efficiency of secondary regularization, since the regularization mesh constrains internal covariate shift and increases limiting convergence deformations. The regularization efficiency coefficient η depends not only on the cross-sectional area of the regularization parameter space but also on the strength characteristics of both the primary optimizer and the dataset scale class.

The study also confirmed the following theoretical findings. The combined use of secondary structured regularization and high-capacity primary optimization substantially increases the predictive strength and gradient deformation resistance of compressed models. The formulas P* and η proposed on the basis of the empirical data enable precise capacity planning in contemporary intelligent system design. Secondary mesh regularization ensures model structural integrity under both short-term evaluation and eccentric (imbalanced) loading conditions.

The study results may be applied in the design and deployment of machine learning pipelines in the following directions: selection of primary optimizer class and secondary regularization type in intelligent system architecture design, and calculation of regularization efficiency; design and fabrication of new types of adaptive neural pipeline elements with an improved accuracy-to-resource-consumption ratio; and assurance of long-term model reliability through the selection of optimal optimizer and regularization combinations for IoT and cyber-physical deployment.

On the basis of the results obtained, the following research directions are recommended: extensive investigation of the effects of different regularization mesh configurations and training epoch pitches; investigation of model resistance to adversarial and distributional shift loads through the combined use of high-capacity primary optimization and modified dataset augmentation strategies; practical verification of the results through full-scale deployment testing on industrial IoT monitoring platforms; and integration of empirical data into automated design methods using AutoML and neural architecture search (NAS) techniques.

All objectives set in the course of this study were fully accomplished. The empirical data and analytical results confirmed the research hypothesis: the combined application of secondary mesh regularization and high-capacity primary optimization increases the predictive strength of compressed models, reduces gradient deformation, and improves the plastic properties of gradient flow. The research results contribute new knowledge of scientific and practical importance to the field of intelligent information system design and machine learning pipeline optimization.

References

Abdullayev, 2023 – Abdullayev, R.K. (2023). Regularization in Graph Neural Networks for Cyber-Physical System Monitoring. Journal of Intelligent Information Systems. Vol. 61. No. 3. Pp. 33–47.

Akhmetov, 2023 – Akhmetov, A.S. (2023). AutoML Pipeline Optimization with Bayesian Regularization. In: Proceedings of the International Conference on Big Data Analytics (ICBDA). Almaty: KazGASA. Pp. 21–35. [In Russ.]

Chen & Liu, 2019 – Chen, W., Liu, Y. (2019). Batch Normalization and SGD Momentum Effects on CNN Convergence under CIFAR-10. Pattern Recognition Letters. Vol. 126. Pp. 45–56.

Goodfellow et al., 2016 – Goodfellow, I., Bengio, Y., Courville, A. (2016). Deep Learning. Cambridge: MIT Press. Pp. 5–17. [In Eng.]

Ioffe & Szegedy, 2015 – Ioffe, S., Szegedy, C. (2015). Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In: Proceedings of ICML 2015. PMLR. Vol. 37. Pp. 448–456.

Kovalev & Smirnov, 2021 – Kovalev, I.A., Smirnov, D.V. (2021). Combined L1/L2 and Dropout Regularization for Industrial Sensor Classification. Vestnik KazGASA. No. 3-4(45-46). Pp. 15–23. [In Russ.]

LeCun et al., 2020 – LeCun, Y., Bottou, L., Orr, G.B., Müller, K.-R. (2020). Efficient Backprop. In: Neural Networks: Tricks of the Trade. Berlin: Springer. Pp. 10–22. [In Eng.]

Nguyen et al., 2020 – Nguyen, T.H., Park, S., Lee, J. (2020). L2 Regularization in LSTM Models for IoT Time-Series Anomaly Detection. Sensors. Vol. 20. No. 8. Pp. 23–32.

Park et al., 2022 – Park, J., Kim, H., Seo, M. (2022). AdamW Weight Decay in Transformer NLP Pipelines: A Systematic Comparison. IEEE Transactions on Neural Networks and Learning Systems. Vol. 33. No. 5. Pp. 45–54.

Zhang & Moldasheva, 2024 – Zhang, Aishuai, Moldasheva, R.N. (2024). Hybrid CNN-LSTM Architectures with Cyclical Learning Rate and Combined Regularization for Intelligent Monitoring Systems. Industrial Transport of Kazakhstan. Vol. 23. No. 1. Pp. 12–28.