https://doi.org/10.51885/3134-8025_JICS_2026_2_7
МРНТИ 28.23.15

# СРАВНИТЕЛЬНЫЙ АНАЛИЗ И ОЦЕНКА ЭФФЕКТИВНОСТИ СВЕРТОЧНЫХ И ТРАНСФОРМЕРНЫХ АРХИТЕКТУР ГЛУБОКОГО ОБУЧЕНИЯ ДЛЯ МНОГОКЛАССОВОЙ КЛАССИФИКАЦИИ МЕДИЦИНСКИХ ИЗОБРАЖЕНИЙ

# МЕДИЦИНАЛЫҚ КЕСКІНДЕРДІ КӨП КЛАСТЫ ЖІКТЕУ ҮШІН ТЕРЕҢ ОҚЫТУДЫҢ КОНВОЛЮЦИЯЛЫҚ ЖӘНЕ ТРАНСФОРМЕРЛІК АРХИТЕКТУРАЛАРЫНЫҢ ТИІМДІЛІГІН САЛЫСТЫРМАЛЫ ТАЛДАУ ЖӘНЕ БАҒАЛАУ

# COMPARATIVE ANALYSIS AND EVALUATION OF THE EFFECTIVENESS OF CONVOLUTIONAL AND TRANSFORMER DEEP LEARNING ARCHITECTURES FOR MULTI-CLASS CLASSIFICATION OF MEDICAL IMAGES

**З.Г. Кабдрахманова[1]*, А.С. Тлебалдинова[1], С.К. Кумаргажанова[1], М.А. Карменова[2], Zbigniew Omiotek[3]**
[1]Восточно-Казахстанский технический университет им. Д. Серикбаева, г. Усть-Каменогорск, Казахстан
[2]Восточно-Казахстанский университет им. С. Аманжолова, г. Усть-Каменогорск, Казахстан
[3]Люблинский технический университет, г. Люблин, Польша
*Автор-корреспондент: Кабдрахманова Зауре Гадыльжановна, e-mail: zkabdrahmanova@vku.edu.kz

**Ключевые слова:**
иерархический трансформер зрения, Swin Transformer, классификация степени ожога, анализ медицинских изображений, глубокое обучение, перенос обучения, компьютерная диагностика, механизмы внимания.

**АННОТАЦИЯ**
Точное определение глубины и площади ожогового поражения – это ключ к правильному выбору метода лечения. Обычно врачи полагаются на визуальную оценку, что зачастую бывает субъективным и может привести к разногласиям между специалистами, а также снизить точность диагностики. Хотя за последние годы активно развиваются методы глубокого обучения для анализа медицинских изображений, потенциал трансформерных моделей в определении степени тяжести ожогов остается недостаточно исследованным.
В нашей работе мы предложили новый подход к автоматической классификации ожогов, основанный на современных архитектурах глубокого обучения. Мы использовали технологии трансферного обучения с предварительно обученными весами, а также применяли аугментацию данных для повышения надежности модели. Для обучения выбрали оптимизатор AdamW и внедрили динамическое управление скоростью обучения, чтобы сделать модель более устойчивой и способной к обобщению. В рамках исследований протестировали разные архитектуры – и сверточные нейронные сети, и трансформеры. Лучший результат показала

© 2026 **З.Г. Кабдрахманова, А.С. Тлебалдинова, С.К. Кумаргажанова, М.А. Карменова, Zbigniew Omiotek**
Данная работа распространяется на условиях лицензии Creative Commons «С указанием авторства» 4.0 Международная (CC BY 4.0).
https://creativecommons.org/licenses/by/4.0/

трансформерная модель, достигшая точности 94,2 % благодаря способности эффективно учитывать как локальные текстурные детали, так и глобальные контекстные признаки.

**Түйінді сөздер:**
иерархиялық көру трансформері, Swin Transformer, күйік дәрежесін жіктеу, медициналық кескінді талдау, терең оқыту, оқытуды тасымалдау, компьютерлік диагностика, назар аудару механизмдері.

**ТҮЙІНДЕМЕ**

Күйіктің тереңдігі мен ауданын дәл анықтау – дұрыс емдеу әдісін таңдаудың кілті. Әдетте дәрігерлер визуалды бағалауға сүйенеді, бұл көбінесе субъективті және мамандар арасындағы келіспеушіліктерге әкелуі мүмкін, сонымен қатар диагностиканың дәлдігін төмендетеді. Медициналық кескіндерді талдау үшін терең оқыту әдістері соңғы жылдары белсенді дамып келе жатқанымен, күйіктердің ауырлығын анықтаудағы трансформаторлық модельдердің әлеуеті әлі зерттелмеген.

Біздің жұмысымызда біз терең оқытудың заманауи архитектураларына негізделген күйіктерді автоматты түрде жіктеудің жаңа әдісін ұсындық. Біз алдын ала дайындалған салмақтармен (веса) трансферлік оқыту технологияларын қолдандық, сонымен қатар модельдің сенімділігін арттыру үшін деректерді күшейтуді қолдандық. Оқыту үшін AdamW оңтайландырушысы таңдалды және модельді тұрақты және жалпылауға қабілетті ету үшін динамикалық оқу жылдамдығын басқаруды енгізді. Зерттеулер шеңберінде әртүрлі архитектуралар – конволюциялық нейрондық желілер де, трансформаторлар да сыналды. Жергілікті текстуралық бөлшектерді де, жаһандық контекстік белгілерді де тиімді есепке алу қабілетінің арқасында 94,2 % дәлдікке жеткен трансформаторлық модель ең жақсы нәтиже көрсетті.

**keywords:**
hierarchical vision transformer, Swin Transformer, burn severity classification, medical image analysis, deep learning, transfer learning, computer diagnostics, attention mechanisms.

**ABSTRACT**

Accurate determination of the depth and area of burn damage is key to choosing the right treatment method. Doctors usually rely on visual assessment, which is often subjective and can lead to disagreements between specialists, as well as reduce the accuracy of diagnosis. Although deep learning methods for medical image analysis have been actively developed in recent years, the potential of transformer models in determining the severity of burns remains understudied.

In our work, we proposed a new approach to automatic burn classification based on modern deep learning architectures. We used transfer learning technologies with pre-trained weights and applied data augmentation to improve model reliability. We chose the AdamW optimizer for training and implemented dynamic learning rate control to make the model more stable and capable of generalization. As part of our research, we tested different architectures, including convolutional neural networks and transformers. The transformer model showed the best results, achieving 94.2% accuracy, thanks to its ability to effectively take into account both local texture details and global contextual features.

## ВВЕДЕНИЕ

Ожоги являются огромной проблемой для здравоохранения многих стран и всего мирового сообщества из-за их распространенности, сложности терапии и значительного влияния на клинические исходы пациентов. Оценка степени и глубины поражения очень важна для выбора правильных лечебных стратегий, прогноза выздоровления и рационального использования медицинских ресурсов. Однако в настоящее время методы оценки во многом, а иногда и в основном, зависят от зрительного осмотра и опыта врача,

вследствие чего наблюдаются значительная индивидуальная субъективность, а также разногласия между разными специалистами по поводу одних и тех же клинических случаев. Различия в опыте врачей, условиях проведения осмотра, а также исключительные для каждого пациента факторы, могут приводить к несогласованному оцениванию и, как следствие, к непредсказуемым ошибкам в ходе принятия лечебных решений. Все вышеизложенное обусловливает потребность в автоматических и объективных системах поддержки принятия решений, которые позволят врачам проводить более последовательную и надежную оценку тяжести ожогов.

Настоящая работа посвящена вопросам применения глубокого обучения (deep learning) и трансферного обучения (transfer learning) для задачи оценки тяжести кожных ожогов. Обсуждаются особенности и результаты каждого из представленных нами алгоритмов, а также их потенциальная роль в практике диагностики ожоговых травм. По нашим результатам методы трансферного обучения могут быть эффективно использованы в диагностике и, соответственно, внедрены в клиническую практику. В последующем это будет важным шагом на пути к развитию новых инструментов, которые позволят своевременно выявлять пациентов за счет совершенствования методов, настройки гиперпараметров и повышения качества данных.

**Обзор литературы.** Развитие искусственного интеллекта и глубокого обучения (deep learning) произвело сдвиг в области анализа медицинских изображений, позволив автоматизировать извлечение признаков и распознавание образов, что превосходит традиционные методы ручного анализа. Особенно успешными стали сверточные нейронные сети (Convolutional Neural Networks, CNN), которые применяются в различных задачах медицинской визуализации. Так, в исследованиях D. P. Yadav и соавт. (2022), а также Priya Mittal и соавт. (2024) были продемонстрированы высокий потенциал и надежность сверточных архитектур в извлечении локальных признаков и текстурных особенностей медицинских изображений, а также их способности обеспечивать высокоточную классификацию и сегментацию за счет иерархического извлечения признаков и увеличения рецептивного поля по мере углубления сети.

В последнее время архитектуры на основе трансформаторов (transformer, vit и vit в частном случае) приобрели большую популярность как эффективный механизм самовнимания (self-attention), способный захватывать как локальные, так и глобальные взаимосвязи в визуальных данных. Одним из таких архитектурных решений является Swin Transformer, представляющий собой иерархическую структуру трансформера с окном внимания со сдвигом. Это позволяет достигать хорошей эффективности вычислений за счет размеров окна, а также поддерживать масштабирумость решения для изображений с высоким разрешением.

В трудах Hu Cao и соавт. (2021) и Kumar Anuj и соавт. (2025) рассматривается применение архитектуры Swin Transformer к анализу медицинских изображений, где одной из главных сложностей является необходимость извлечения высокоразмерных признаков при ограниченном объеме размеченных данных. Модель реализует многозадачное обучение, что позволяет успешно решать несколько разных задач медицинской визуализации, одновременно формируя обобщённые представления, применимые к новым данным и к новым задачам. Благодаря адаптивному механизму внимания (attention), архитектура способна фокусировать общее внимание только на значимых частях изображений, что позволяет снизить эффект шума, артефактов и низкоконтрастных изображений.

Кроме того, в работе предложен повторный encoder с иерархической структурой и многоуровневым анализом, что повышает захват локальных и глобальных сценариев. По сути, эксперименты показали, что эффективность модели достигает точности 80,76%.

Таблица 1 содержит обзор некоторых связанных работ (в таблице указаны используемые методы и достигнутая точность классификации изображений ожогов).

**Таблица 1.** Обзор литературы и оценка методов и моделей глубокого обучения для распознавания ожоговых травм

| Авторы, название, год | Методы и модели распознавания ожоговых поражений | Результаты исследования |
|---|---|---|
| Hassanipour S. et al., Comparison of artificial neural network and logistic regression models for prediction of outcomes in trauma patients: A systematic review and metaanalysis, 2019 | ANN и LR | Уровень точности ANN и LR в моделях случайных эффектов составил 90,5 (95 % CI 87,6–94,2) и 83,2 (95 % CI 75,1–91,2) соответственно |
| Cirillo M. D., Mirdell R., Sjöberg F. et al., Time-independent prediction of burn depth using deep convolutional neural networks, 2019 | CNN: VGG-16, GoogleNet, ResNet-50, и ResNet-101 | ResNet-101: средняя точность, чувствительность и специфичность для четырех видов ожогов составили 90,54 %, 74,35 % и 94,25 % соответственно |
| Uğur Şevik, Erdinç Karakullukçu, Tolga Berber, Yeşim Akbaş, Serdar Türkyılmaz. Automatic classification of skin burn color images using texture-based feature extraction, 2019 | SegNet и CNN | Имея F-оценку 74,28 % при классификации изображений, полученных без протокола, предложенная схема смогла достичь аналогичных результатов с глубоким обучением, которое имело F-оценку 80,50 %. |
| Боб Чжан, Цзяньхан Чжоу Multi-feature representation for burn depth classification using burn images, 2020 | Модель Stacked Sparse Autoencoder (SSAE) применяется для извлечения скрытых признаков (latent features) Random Forest (основная модель для классификации) | точность 85,86 % и 76,87 % путем классификации изображений ожогов на два класса и три класса соответственно, что превосходит обычные методы в определении глубины ожога. Результаты показывают, что данный подход эффективен и может помочь медицинским экспертам в определении различной глубины ожогов. |
| D.P. Yadav, A. S. Jalal, V. Prakash — Human burn depth and grafting prognosis using ResNeXt topology based deep learning network, 2022 | Deep learning модель на основе ResNeXt topology; CNN с residual connections для классификации глубины ожога и прогноза grafting | Deep learning модель на основе ResNeXt topology; CNN с residual connections для классификации глубины ожога и прогноза grafting. (Название модели BNeXtmodel, точность 97,17 %) |
| Ze Liu et al. — Swin Transformer: Hierarchical Vision Transformer using Shifted Windows, 2021 | Hierarchical Swin Transformer с механизмом shifted window attention | Top-1 Accuracy = 0,873 (ImageNet benchmark) |
| Hu Cao et al. — Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation, 2021 | Transformer-based encoder-decoder архитектура (Swin-Unet) для медицинских изображений | Transformer-based encoder-decoder architecture (Swin-Unet) |
| Dosovitskiy et al. — An Image is Worth 16×16 Words: Transformers for Image Recognition at Scale (Vision Transformer), 2021 | Vision Transformer (ViT); чистая transformer архитектура для классификации изображений | Top-1 Accuracy = 0,884 (ImageNet pretraining with large-scale data) |

*Примечание – составлено авторами*

**МАТЕРИАЛЫ И МЕТОДЫ ИССЛЕДОВАНИЯ**

Предлагаемая в данной работе структура глубокого обучения базируется на предварительно обученной иерархической архитектуре трансформатора зрения и позволяет автоматически классифицировать степень тяжести ожоговых травм. Методология также объединяет перенос обучения, структурированное увеличение количества данных, адаптивные стратегии оптимизации и сложное оценивание с целью обеспечения надежности и обобщающей способности при работе с медицинскими данными. Всю структуру данной работы можно представить в виде следующей последовательности: предварительная обработка данных; извлечение характеристик; обучение с учителем; выбор модели, основываясь на валидации; оценка, основываясь на интерпретируемости. Входной датасет состоит из RGB-изображений, сгруппированных на три класса ожогов. Данные разделены на тренировочные изображения – 70%, валидации – 15% и тестирования – 15% в обучении с учителем. Все изображения были форматированы для приведения к фиксированному двумерному разрешению, чтобы их можно было встраивать в механизм патча основного трансформатора. Так как использовались предобученные модели, обученные на ImageNet, входные изображения были нормализованы с использованием средних и стандартных отклонений набора данных для выравнивания распределения функций между предобученной и медицинской областями. Чтобы повысить обобщение и содействовать противопереобучению, вызванному ограниченностью медицинских данных, во время обучения использовалась стохастическая аугментация данных. Эта стратегия приводит к геометрическим аугментациям, таким как отображение по горизонтали и вращение, а также фотометрическим аугментациям, таким как яркость, контраст и насыщенность. Формально, если x – это наше исходное изображение, T – некоторое случайное преобразование, выбранное из множества преобразований T~T, а цель обучения задается на основе преобразованных вариантов изображения $\tilde{x} = T(x)$, то возможности разнообразия используемого для обучения набора расширяются.

Ядром модели для извлечения признаков является Swin Transformer иерархическая архитектура трансформатора зрения (vision transformer), которая внедряет локализованное self-attention в сдвинутых оконных областях. В отличие от классических трансформаторов Vision Transformer (ViT), вычисляющих глобальное self-attention между всеми токенами, Swin Transformer ограничивает внимание неперекрывающимися окнами и периодически сдвигает их, чтобы обеспечить взаимодействие между соседними областями изображения. Такой подход снижает вычислительную сложность, сохраняя при этом моделирование долгосрочной зависимости. Входное изображение разбивается на неперекрывающиеся патчи размером P×P, формируя последовательность вложений токенов:

$$\mathrm{Z}_0 = \mathrm{Embed}(X) \in \mathbb{R}^{N \times D}, \quad (1)$$

где $N = \frac{HW}{P^2}$ обозначает количество патчей, $D$ – размерность embedding space.

В каждом блоке Swin вычисляется оконное multi-head self-attention, при котором формируются запросы (Q), ключи (K) и значения (V):

$$\mathrm{Q} = \mathrm{ZW}_Q, \mathrm{K} = \mathrm{ZW}_K, \mathrm{V} = \mathrm{ZW}_V, \quad (2)$$

операция self-attention внутри локального окна определяется следующим образом:

$$\mathrm{Attn}(\mathrm{Q}, \mathrm{K}, \mathrm{V}) = \mathrm{Softmax}\left(\frac{\mathrm{QK}^{\mathrm{T}}}{\sqrt{d}} + \mathrm{B}\right)\mathrm{V}, \quad (3)$$

где d – размерность головы внимания, B – относительное позиционное смещение (relative position bias). Каждый блок включает residual connections и feed-forward network (MLP):

$$\mathrm{Z}' = \mathrm{Z} + \mathrm{W} - \mathrm{MSA}(\mathrm{LN}(\mathrm{Z})),$$
$$\mathrm{Z}_{out} = \mathrm{Z}' + \mathrm{MLP}(\mathrm{LN}(\mathrm{Z}')). \quad (4)$$

Иерархическое представление строится с помощью операций объединения патчей, которые уменьшают разрешение по пространству и увеличивают размерность канала, таким образом формируя пирамиду признаков, подобную сверточным нейронным сетям. Головка классификации использует глобальное усреднение и линейный слой:

$$o = W_{cls}g + b_{cls}, \tag{5}$$

где g является агрегированным представлением признаков. Вероятности классов вычисляются с помощью softmax:

$$\hat{y}_c = \frac{e^{o_c}}{\sum_{k=1}^{C} e^{o_k}}. \tag{6}$$

Целью оптимизации служит категориальная кросс-энтропийная ошибка:

$$\mathcal{L}(\theta) = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{i,c}\log\hat{y}_{i,c}, \tag{7}$$

где $\theta$ обозначает параметры модели. Обучение производится с помощью оптимизатора AdamW, который разделяет затухание весов и обновления на основе градиента:

$$\theta_{t+1} = \theta_t - \eta_t\frac{\hat{m}_t}{\sqrt{\hat{v}_t}+\epsilon} - \eta_t\lambda\theta_t. \tag{8}$$

Изменение скорости обучения реализовано с помощью ReduceLROnPlateau, при котором темп снижается, если валидационные потери перестают уменьшаться, что позволяет «переключаться» от грубой к более тонкой настройке.

Вычислительная сложность классического глобального самоуправления непомерна для изображений с высоким разрешением. Swin Transformer снижает сложность до:

$$O(N \cdot M^2), \tag{9}$$

где $M$ обозначает размер окна. Эта стратегия локального внимания значительно улучшает масштабируемость, сохраняя при этом выразительную силу.

Были внесены определённые изменения, направленные на оптимизацию результатов при работе с медицинскими изображениями. Во-первых, перенос обучения с использованием весов, предварительно обученных на ImageNet, даёт возможность эффективно захватывать визуальные признаки низкого и среднего уровней, что способствует быстрому обучению и стабильности при недостаточности данных. Во-вторых, аугментация медицинскими данными обеспечивает инвариантность к нейтральным к диагнозу преобразованиям, сохраняя при этом изменения, специфичные для патологии. В-третьих, адаптивное управление оптимизацией с помощью AdamW совместно с регулировкой скорости обучения на основе результатов валидации стабилизирует процесс обучения. В-четвёртых, выбор модели на основе её точности на валидационном наборе обеспечивает способность к обобщению, а не к переобучению на обучающем наборе. И наконец, использование различных метрик, в том числе матриц погрешностей и мультиклассовых метрик, даёт возможность оценивать модель с учётом различных аспектов, что необходимо для её внедрения в клиническую практику. Оптимизированный процесс обучения отличается от классического Swin Transformer в основном стратегиями оптимизации и предварительной обработки данных, но не архитектурными изменениями. Эти улучшения настраивают модель на специфические особенности медицинских изображений, что и даёт высокий результат при задачах классификации. Возможность Swin Transformer строить иерархические представления, а также использовать локальное внимание, позволяет эффективно улавливать как детали текстур, так и широкие контексты, влияющие на содержание снимков при оценке степени ожогов. Такой подход особенно эффективен при работе с медицинскими текстурами, что и объясняет высокий результат

данного метода на задаче, на которую он был применён. На рисунке 1 показана архитектура Swin трансформера для классификации ожогов.

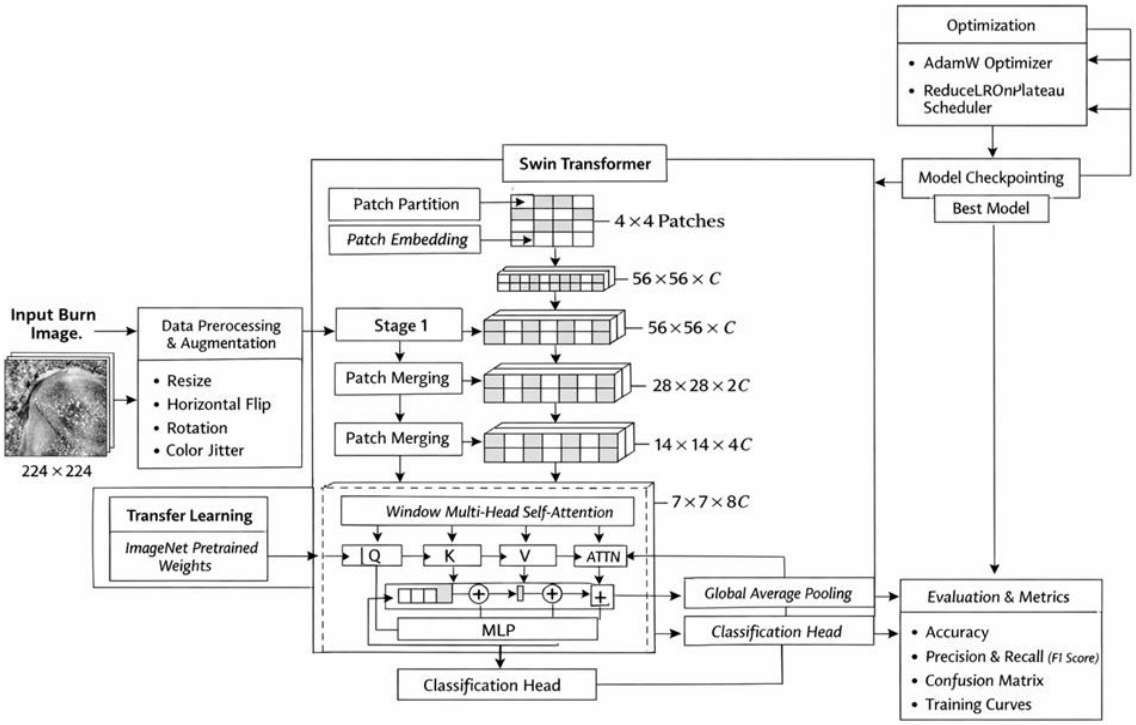



**Рисунок 1.** Архитектура Swin трансформерной модели, предлагаемая для классификации ожогов кожи

*Примечание – составлено авторами*

**Реализация алгоритма.** Модель классификации ожогов по степени тяжести, предложенная на базе Swin Transformer, была реализована с помощью PyTorch фреймворка глубокого обучения и библиотеки моделей Timm для трансформерных моделей зрения. Разработка и тесты проводились в научной вычислительной среде Spyder, что позволило сделать процессы обучения, оценки и визуализации воспроизводимыми. Данные были разделены на три поднабора для обучения, валидации и тестирования, соответственно, при этом изображения были классифицированы на три категории степени тяжести ожогов. Все входные изображения были приведены к фиксированному разрешению 224 × 224 пикселей в соответствии с требованиями архитектуры базовой структуры Swin Transformer. Перед обучением изображения были нормализованы на основе статистики ImageNet, чтобы достичь совместимости с предварительно обученными весами модели. Для увеличения обобщающей способности модели и снижения переобучения, что является распространенной проблемой в задачах медицинской визуализации, в процессе обучения использовались техники расширения данных. Среди них были горизонтальное отражение, случайное вращение и преобразования цветового дрожания, которые добавляют небольшую вариативность, сохраняя при этом важные для лечения признаки визуализаций. В этом исследовании была использована базовая модель Swin Transformer Tiny (Swin_Tiny_Patch4_Window7_224), которая была инициализирована весами, предобученными на ImageNet, для применения переноса обучения. Классификационная головка была автоматически перенастроена в соответствии с количеством выходных классов. Оптимизация модели была выполнена с помощью оптимизатора AdamW, который различает затухание весов и обновления градиента для улучшения стабильности сходимости при работе с трансформерными

архитектурами. Обучение модели производилось в течение 100 эпох при размере батча 8 и начальной скорости обучения 2, $0{\times}10^{-4}$ и коэффициентом затухания веса 1, $0{\times}10^{-4}$. В качестве функции ошибки при обучении использовалась категориальная кросс-энтропия. Для динамического корректирования значения скорости обучения применялась стратегия планирования ReduceLROnPlateau, которая уменьшает скорость обучения при отсутствии улучшений значения функции потерь на валидационном наборе. Обновление значения скорости обучения может быть выражено следующим уравнением:

$$\eta_{t+1} = \begin{cases} \gamma \cdot \eta_t, & \text{if } \mathcal{L}_{val} \quad \text{не улучшается} \\ \eta_t, & \text{иначе ,} \end{cases} \tag{10}$$

где $\eta_t$ обозначает скорость обучения при итерации $t$, $\gamma -$ коэффициент затухания (установлен на 0,3), а $\mathcal{L}_{val}$ представляет собой потерю валидации.

Во время обучения по мере улучшения метрик обучения и валидации, таких как точность и потери, лучшая модель сохранялась на диск. На этапе тестирования были собраны и рассчитаны значения всех метрик с accuracy, precision, recall, F1-score и построена матрица неточностей для более глубокого анализа поведения модели в отношении разных классов. Все эксперименты и тренировки были проведены на системе с процессорами Intel(R) Core (TM) i5-12500H 12-го поколения с тактовой частотой 2, 50 ГГц, 8 ГБ оперативной памяти и 64-разрядной операционной системой x64.

**РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ**

Результаты экспериментов показывают, что при учёте и локальной текстуры, и глобальных контекстных связей в медицинских изображениях архитектуры на базе трансформеров оказываются эффективнее, чем традиционные свёрточные архитектуры. В эксперименте использовались определенные конфигурации моделей и стратегия обучения; основные гиперпараметры, использованные в данной работе, приведены в табл. 2.

**Таблица 2.** Конфигурация модели и гиперпараметры обучения для архитектуры Swin Transformer

| Категория | Параметр | Значение | Описание |
|---|---|---|---|
| Конфигурация архитектуры | Базовая модель (Backbone) | swin_tiny | Hierarchical vision transformer with shifted windows attention |
| | Размер входного изображения | 224 × 224 | Стандартизованный размер входного изображения |
| | Размер патча | 4 × 4 | Размер патча для разбиения изображения |
| | Размер окна | 7 × 7 | Размер окна локального self-attention |
| | Размер embedding | 96 | Начальный размер embedding |
| | Количество иерархических стадий | 4 | Глубина иерархической обработки |
| | Предобучение | weights ImageNet | Использование предобученной модели на ImageNet |
| Параметры оптимизации | Оптимизатор | AdamW | Используемый оптимизатор |
| | Начальная скорость обучения | $2 \times 10^{-4}$ | Начальная скорость обучения |
| | Weight decay | $1 \times 10^{-4}$ | Коэффициент регуляризации |
| | Декремент скорости обучения | ReduceLROnPlateau | Снижение скорости обучения при отсутствии улучшении метрики |

Окончание таблицы 2

| Категория | Параметр | Значение | Описание |
|---|---|---|---|
| | Коэффициент снижения LR | 0.3 | Коэффициент уменьшения learning rate |
| | Минимальный learning rate | $1 \times 10^{-6}$ | Нижняя граница learning rate |
| Настройки обучения | Размер batch | 8 | Размер batch при обучении |
| | Количество эпох | 100 | Максимальное количество эпох при обучении |
| | Функция потерь | CrossEntropyLoss | Тип функции потерь |
| | Random seed | 42 | Случайное зерно для воспроизводимости результатов |
| Процедура предобработки данных | Resize | 224 × 224 | Изменение размера изображений |
| | Normalize | mean/std of ImageNet | Используются те же средние значения и стандарты отклонения, что и при предобучении на ImageNet |
| Аугментация данных | Горизонтальное отражение | Включено | Для увеличения объема тренировочных данных |
| | Случайное вращение | ± 10° | Для увеличения объема тренировочных данных |
| | Color jitter | Brightness (0,15), Contrast (0,15), Saturation (0,1), Hue (0,02) | Для увеличения объема тренировочных данных |
| Оценка | Метрики | Accuracy, Precision, Recall, F1-score | Множество метрик для оценки качества классификации |
| | Матрица ошибок | Используется | Для изучения характера ошибок классификатора |

*Примечание – составлено авторами*

Для оценки надежности и воспроизводимости результатов модели их модель была обучена 3 раза с различными случайными инициализациями и разными random seed. Результаты представлены в виде средних значений с соответствующими стандартными отклонениями. Низкая изменчивость при повторных запусках указывает на стабильность и устойчивость модели к случайным факторам обучения.

Модели, обученные в этом эксперименте, были оценены с использованием количественных метрик производительности для классификационной эффективности по всем категориям. Результаты экспериментов предоставляют сравнительный обзор для выделения моделей, улучшая при этом набор критериев точности классификации, precision, recall и F1-score. Таким образом, можно более объективно оценить, насколько различные архитектурные решения влияют на предсказательную эффективность в рамках решения конкретной задачи. Accuracy - это процент объектов, по которым правильно предсказана их категория из общего числа примеров. В задачах многоклассовой классификации показатель accuracy (ACC) вычисляется так же, как и в бинарном случае и отражает долю правильно классифицированных образцов от общего числа наблюдений. При этом каждый образец вносит равный вклад в итоговое значение метрики. Такой показатель позволяет оценить, как справляется модель в целом, а также получить представление о её поведении по отдельным классам. Особенно актуально это, если классы несбалансированы или ошибки классификации имеют разную важность.

$$\text{Accuracy: } ACC = \frac{TP+TN}{TP+TN+FP+FN}, \quad (11)$$

где: TP – количество правильно предсказанных положительных результатов; TN — количество правильно предсказанных отрицательных результатов; FP — количество неправильно предсказанных положительных результатов; FN — количество неправильно предсказанных отрицательных результатов.

Точность определяется как доля объектов, которые классифицированы как положительные и действительно являются положительными. В задаче многоклассовой классификации объекты классифицируются более чем по двум классам (степени ожогов: легкая, средняя, тяжелая). Для каждого класса k метрики рассчитываются отдельно в соответствии с принципом «один против всех». Точность для класса k определяется как отношение числа объектов, корректно отнесённых к классу k, к общему количеству объектов, предсказанных моделью как принадлежащих классу k:

$$\text{Precision: } PRE_k = \frac{TP_k}{TP_k + FP_k}, \tag{12}$$

Возврат для класса k – это доля объектов в классе k, которые были правильно распознаны.

$$\text{Recall: } REC_k = \frac{TP_k}{TP_k + FN_k}, \tag{13}$$

F1 – показатель для класса k, является средним гармоническим показателей точности и полноты для класса k:

$$\text{F1- score: } F1_k = 2\frac{PRE_k \cdot REC_k}{PRE_k + REC_k}. \tag{14}$$

Числовые результаты экспериментов приведены в следующей таблице.

**Таблица 3.** Сравнение результатов использования различных моделей глубокого обучения для задачи классификации ожогов кожи

| № | Название модели | Accuracy | Precision | Recall | F1-score |
|---|---|---|---|---|---|
| 1 | Xception | 0.7565 | 0.7607 | 0.7562 | 0.7468 |
| 2 | VGG16 | 0.9188 | 0.9193 | 0.9187 | 0.9189 |
| 3 | VGG19 | 0.9041 | 0.9053 | 0.9040 | 0.9044 |
| 4 | ResNet50 | 0.8967 | 0.8992 | 0.8966 | 0.8973 |
| 5 | DenseNet201 | 0.8708 | 0.8769 | 0.8708 | 0.8723 |
| 6 | MobileNetV2 | 0.8469 | 0.8467 | 0.8468 | 0.8459 |
| 7 | Inception_resnet_v2 | 0.8704 | 0.8804 | 0.8704 | 0.8701 |
| 8 | EffNetB0 | 0.7915 | 0.7929 | 0.7911 | 0.7903 |
| 9 | AttnMS_VGG16 | 0.9041 | 0.9052 | 0.9040 | 0.9044 |
| 10 | Vgg16_TTA | 0.9004 | 0.9053 | 0.9004 | 0.9015 |
| 11 | Vit_TTA | 0.9059 | 0.9071 | 0.9058 | 0.9063 |
| 12 | Convnext_tiny | 0.7897 | 0.8018 | 0.7897 | 0.7902 |
| 13 | Swin_tiny | 0.9428 | 0.9426 | 0.9427 | 0.9425 |
| *Примечание – составлено авторами* | | | | | |

Сравнительный анализ показывает, что архитектуры, основанные на трансформерах, лучше традиционных сверточных нейронных сетей справляются с классификацией степени тяжести ожогов. Из всех моделей наилучшие результаты были у иерархического Swin Transformer, что подтверждает его способность эффективно улавливать как локальные текстуры, так и общие контекстные связи в медицинских изображениях. Классические архитектуры CNN, такие как VGG16 и VGG19, показали хорошие базовые результаты и подтвердили свою эффективность в задачах классификации на основе текстуры в медицине, однако они были слегка уступлены моделям на трансформерах.

Легкие и ориентированные на эффективность модели, такие как MobileNetV2 и ConvNext_tiny, показали сравнительно низкую производительность, что, вероятно, связано с тем, что снижение сложности модели ограничивает ее возможности представить данные. Варианты с улучшенным вниманием и стратегия увеличения времени тестирования повысили стабильность и надежность, что еще раз подчеркивает важность применения передовых методов для характеристики моделей.

Качество моделей классификации оценивалось с помощью теста МакНемара, применяемого для сравнения парных прогнозов на одном и том же тестовом наборе данных. В частности, были сравнены модели Swin Transformer и VGG16, показавшие максимальную точность. Тест был проведен на основе анализа несовпадающих пар прогнозов. Статистическая значимость признавалась при $p < 0,05$.

Ниже представлена сравнительная диаграмма оценок метрик для каждой модели.

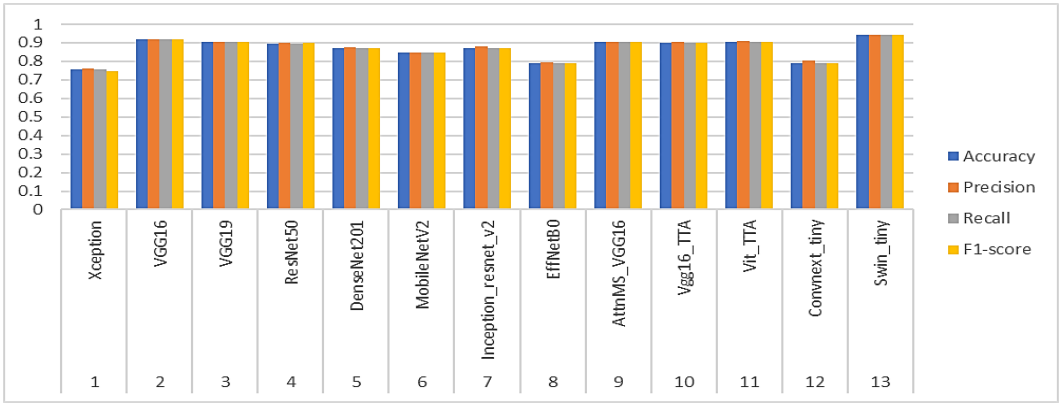


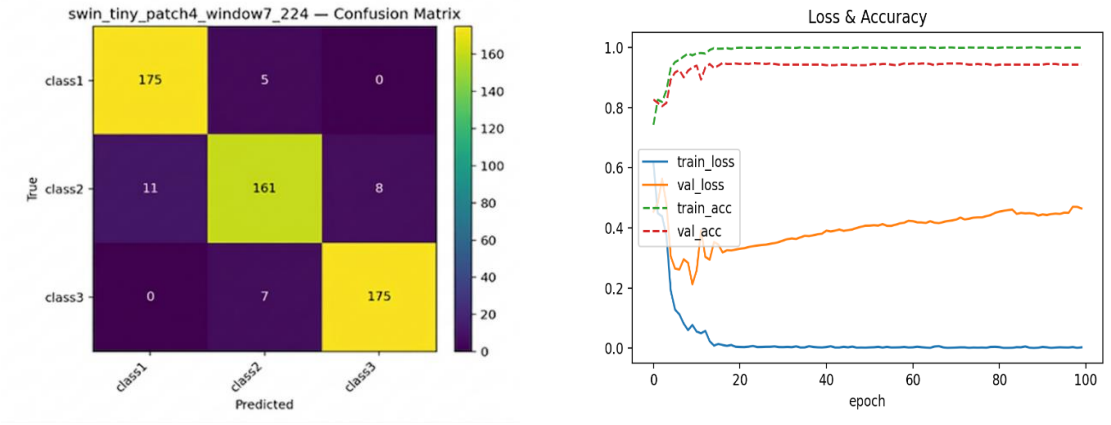

**Рисунок 2.** Сравнительная диаграмма показателей

*Примечание – составлено авторами*

Результаты, показанные на рисунке 3 (а), подтверждаются и диаграммой ошибок, на которой видно, что классификация проведена качественно, о чем свидетельствуют многочисленные правильные прогнозы на главной диагонали для всех трех классов. Особенно хорошо выявляются классы 1 и 3: они с высокой вероятностью определяются верно, незначительно перепутываются, между ними не происходит прямого замешательства, что говорит о хорошей отделимости их признаков. Большинство ошибок связаны со вторым классом: он перепутан с первым, порой с третьим, что указывает на то, что в признаках этого класса или в промежуточных признаках соседних классов есть что-то общее. Несмотря на эти проблемы, общее количество ошибок невелико, что говорит о том, что модель способна обобщать знания по сбалансированным классам, правильно выделяя как характерные особенности, так и контекст. Однако до сих пор модель не может хорошо отделить второй класс от двух других. В целом же по матрице ошибок можно сказать, что модель справляется с нечёткими ситуациями и делает сбалансированный выбор, что говорит о возможности использовать ее для достаточно надежной классификации медицинских изображений в нескольких классах.

Рисунок 3 (б) показывает кривые обучения, по которым мы можем следить за изменением значений функций потерь и достигаемых точностей в процессе обучения с возрастанием числа эпох. Потери при обучении уменьшается очень быстро, приближаясь к нулю в течение нескольких эпох и тем самым доказывая, что модель успешно адаптируется под обучающую выборку.

Одновременно с этим точность при обучении тоже быстро растёт и достигает очень высокого значения, демонстрируя высокую способность модели к обучению.




а) б)

**Рисунок 3.** Матрица ошибок и loss модели Swin tiny

*Примечание – составлено на основе расчетов авторов*

**ЗАКЛЮЧЕНИЕ**

В данной статье представлены результаты экспериментов с использованием глубокого обучения при автоматическом определении степени тяжести ожогов на основе медицинских изображений, собранных из общедоступных источников. Сбалансированный набор из трех классов, представляющих разные степени тяжести ожогов, был собран из открытых репозиториев, таких как Kaggle и Roboflow. Для каждого класса было отобрано ровно 1200 образцов, из них около 840-848 изображений на класс были использованы для обучения, остальные изображения были отложены для валидации и тестирования модели. Благодаря сбалансированной структуре набора данных результаты модели можно оценивать объективно. Данный баланс минимизирует проблему смещения модели из-за дисбаланса классов во время обучения.

Были протестированы тринадцать архитектур глубокого обучения. Рассматривались классические сверточные нейронные сети (VGG16, VGG19, ResNet50, DenseNet201, MobileNetV2, Xception и Inception-based модели) и новые сверточные архитектуры (ConvNeXt), а также модель с улучшенным вниманием и модели на основе трансформеров (Vision Transformer (ViT) и Swin Transformer). Такая работа позволила детально проанализировать, как различные архитектурные схемы влияют на качество решения задачи классификации степени ожога по медицинским изображениям. Для обеспечения эффективности и стабильности модели был внедрен ряд стратегий и методов оптимизации. В том числе: перенос обучения, при котором используются веса предварительно обученной модели; увеличение объема данных с учетом домена; планирование скорости обучения с теплым перезапуском; улучшение характеристик с помощью внимания; увеличение объема данных во время тестирования для улучшения способности к обобщению.

Эффективность моделей также оценивалась с помощью комплексных показателей и метрик, таких как общая точность, точность, воспроизводимость, показатель F1 и анализ матрицы путаницы. Результаты экспериментов показывают, что учет как локальной текстуры, так и глобальных контекстуальных отношений в медицинских изображениях с помощью архитектур на основе трансформаторов превосходит по эффективности традиционные архитектуры на основе свертки. Вклад данной статьи заключается в разработке интеллектуальных систем поддержки принятия клинических решений путем предоставления структурированного подхода к оценке современных архитектур глубокого

обучения для классификации глубины ожогов. Новизна работы заключается в том, что в ней исследуется сравнительная эффективность различных глубоких архитектур, включая CNN и модели на основе внимания, для классификации степени повреждения с использованием цветных медицинских изображений. Были предложены соответствующие стратегии обучения для оптимизации производительности для конкретных задач, включая настройку гиперпараметров, аугментацию данных, интеграцию моделей на основе трансформаторов, реализованных в сочетании с оптимизированными стратегиями обучения, иерархические механизмы внимания, адаптированные к области задачи анализа ожогов. Результаты, полученные в этой работе, являются алгоритмической основой и научно-методическим руководством для последующей разработки информационной системы для поддержки диагностики и лечения ожогов.

В итоге результаты показывают, что архитектуры на основе трансформеров являются многообещающим направлением для повышения точности, объективности и надежности автоматической классификации степени ожогов и, несомненно, будут стимулировать будущие исследования в области создания клинически применимых систем искусственного интеллекта.

**КОНФЛИКТ ИНТЕРЕСОВ:** Авторы заявляют об отсутствии конфликта интересов.

**ФИНАНСИРОВАНИЕ:** Работа выполнена при поддержке финансирования проекта №AP23486396 «Модели и методы распознавания анатомических структур на изображениях МРТ в задачах компьютерной диагностики» на 2024–2026 годы Министерства науки и высшего образования Республики Казахстан.

**ЗАЯВЛЕНИЕ ОБ ОДОБРЕНИИ ИНСТИТУЦИОНАЛЬНЫМ ЭТИЧЕСКИМ КОМИТЕТОМ (IRB):** Данное исследование имеет исключительно техническую направленность и не предусматривает участие людей, животных или проведение биологических экспериментов, поскольку все эксперименты выполняются исключительно на изображениях ожоговых ран, в связи с этим данный раздел отмечен как «не применимо»

**ЗАЯВЛЕНИЕ ОБ ИНФОРМИРОВАННОМ СОГЛАСИИ:** *Так как все эксперименты выполняются исключительно на изображениях ожоговых поражений, данный раздел считается «не применимым»*

**ЗАЯВЛЕНИЕ О ДОСТУПНОСТИ ДАННЫХ:** Данные, использованные в этом исследовании, были получены из открытых источников, таких как Kaggle и Roboflow, и находятся в свободном доступе для использования. Все наборы данных могут быть загружены и использованы без ограничений. В рамках данного исследования новые экспериментальные данные не создавались.

**БЛАГОДАРНОСТИ:** Авторы выражают признательность коллегам за методологическую поддержку и конструктивные обсуждения, а также анонимным рецензентам за ценные замечания, способствовавшие повышению качества представленной работы.

**УВЕДОМЛЕНИЕ ОБ ИСПОЛЬЗОВАНИИ ТЕХНОЛОГИЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА:** В данной публикации не использовались инструменты искусственного интеллекта (ИИ)

**СПИСОК ЛИТЕРАТУРЫ**

Rozo, A., Miskovic, V., Rose, T., Keersebilck, E., Iorio, C., & Varon, C. (2023). A deep learning image-to-image translation approach for a more accessible estimator of the healing time of burns. Artificial Intelligence in Medicine, 140. https://doi.org/10.1016/j.artmed.2023.102570

Yadav, D. P., Jalal, A. S., & Prakash, V. (2022). Human burn depth and grafting prognosis using

ResNeXt topology based deep learning network. Multimedia Tools and Applications, 81(13), 18897-18914. https://doi.org/10.1007/s11042-022-12446-0

Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. In International Conference on Learning Representations (ICLR). https://doi.org/10.48550/arXiv.1409.1556

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR). https://doi.org/10.1109/CVPR.2016.90

Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR). https://doi.org/10.1109/CVPR.2017.195

Huang, G., Liu, Z., van der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR). https://doi.org/10.1109/CVPR.2017.243

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2018). MobileNetV2: Inverted residuals and linear bottlenecks. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR). https://doi.org/10.1109/CVPR.2018.00474

Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. (2017). Inception-v4 Inception-ResNet and the impact of residual connections on learning. In AAAI Conference on Artificial Intelligence. https://doi.org/10.1609/aaai.v31i1.11231

Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. In International Conference on Machine Learning (ICML). https://doi.org/10.48550/arXiv.1905.11946

Dosovitskiy, A., et al. (2021). An image is worth 16×16 words: Transformers for image recognition at scale. In International Conference on Learning Representations (ICLR). https://doi.org/10.48550/arXiv.2010.11929

Liu, Z., et al. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. In IEEE International Conference on Computer Vision (ICCV). https://doi.org/10.1109/ICCV48922.2021.00986

Liu, Z., et al. (2022). ConvNeXt: A ConvNet for the 2020s. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR). https://doi.org/10.1109/CVPR52688.2022.01167

Zhang, B., & Zhou, J. (2021). Multi-feature representation for burn depth classification using burn images. Artificial Intelligence Review. https://doi.org/10.1007/s10462-021-10055-6

Kumar, A. (2025). An improved feature extraction algorithm for robust Swin transformer model in high-dimensional medical image analysis. Computers in Biology and Medicine, 188. https://doi.org/ 10.1016/j.compbiomed.2025.109822

**Авторлар туралы мәліметтер**
**Информация об авторах**
**Information about authors**

**Кабдрахманова Зауре Гадыльжановна** – «Ақпараттық жүйелер» білім беру бағдарламасының докторанты, Д. Серікбаев атындағы Шығыс Қазақстан техникалық университеті, Өскемен қ., Қазақстан.

**Кабдрахманова Зауре Гадыльжановна** – докторант образовательной программы «Информационные системы», Восточно-Казахстанский технический университет им. Д. Серикбаева, г. Усть-Каменогорск, Казахстан.

**Kabdrakhmanova Zaure Gadylzhanovna** – doctoral student of the educational program "Information Systems", D. Serikbaev East Kazakhstan Technical University, Ust-Kamenogorsk, Kazakhstan,
e-mail: zkabdrahmanova@vku.edu.kz,
ORCID: https://orcid.org/0000-0003-4763-9437

**Тлебалдинова Айжан Солтангалиевна** – PhD, қауымдастырылған профессор, *Д*. Серікбаев атындағы Шығыс Қазақстан техникалық университеті, Өскемен, Қазақстан.
**Тлебалдинова Айжан Солтангалиевна** – PhD, ассоциированный профессор, Восточно-Казахстанский технический университет им. *Д*. Серикбаева, Усть-Каменогорск, Казахстан.
**Tlebaldinova Aizhan** – PhD, associate professor, D. Serikbaev East Kazakhstan Technical University, Ust-Kamenogorsk, Kazakhstan
e-mail: atlebaldinova@edu.ektu.kz,
ORCID: https://orcid.org/0000-0003-1271-0352

**Кумаргажанова Сауле Кумаргажановна** – техникалық ғылымдар кандидаты, профессор, *Д*. Серікбаев атындағы Шығыс Қазақстан техникалық университеті, Өскемен қ., Қазақстан.
**Кумаргажанова Сауле Кумаргажановна** – кандидат технических наук, профессор, Восточно-Казахстанский технический университет им. *Д*. Серикбаева, г. Усть-Каменогорск, Казахстан.
**Kumargazhanova Saule Kumargazhanovna** – candidate of Technical Sciences, professor, D. Serikbayev East Kazakhstan Technical University, Oskemen, Kazakhstan,
E-mail: skumargazhanova@gmail.com,
ORCID: https://orcid.org/0000-0002-6744-4023

**Карменова Мархаба Ахметоллиновна** – PhD, қауымдастырылған профессор, С.Аманжолов атындағы Шығыс Қазақстан университеті, Өскемен қ., Қазақстан.
**Карменова Мархаба Ахметоллиновна –** PhD, ассоцированный профессор, Восточно-Казахстанский университет им. С. Аманжолова, г. Усть-Каменогорск, Казахстан.
**Karmenova Markhaba** – PhD, Associate Professor, S. Amanzholov East Kazakhstan University, Ust-Kamenogorsk, Kazakhstan.
e-mail: mkarmenova@vku.edu.kz,
ORCID: https://orcid.org/0000-0003-1271-0352

**Zbigniew Omiotek** – қауымдастырылған профессор, *Люблин* технологиялық университеті, *Люблин*, Польша.
**Zbigniew Omiotek** – ассоцированный профессор, *Люблинский* Технологический Университет, *Люблин*, Польша.
**Zbigniew Omiotek** – associate Professor, Lublin University of Technology, Lublin, Poland,
e-mail: z.omiotek@pollub.pl,
ORCID: https://orcid.org/0000-0002-6614-7799