

## ШВИДКОДІЮЧІ МОДЕЛІ ШТУЧНОГО ІНТЕЛЕКТУ ДЛЯ РЕАЛЬНОГО ЧАСУ В ЦИФРОВИХ ДВІЙНИКАХ ПРОЦЕСІВ ОБРОБКИ МЕТАЛІВ ТИСКОМ

У статті представлено комплексну методичку створення цифрового двійника технологічних процесів обробки металів тиском, який поєднує класичне моделювання методом скінченних елементів і високошвидкісні моделі штучного інтелекту. Метою дослідження було забезпечити інференс із затримкою не більше ніж сто мілісекунд, прискорити повний цикл симуляції утричі з наляком, але в реальності в понад п'ятдесят разів порівняно з традиційними FEM-розрахунками, а також досягти прогнозуної достовірності вище дев'яноста відсотків. Для цього було сформовано навчальний набір із понад дванадцяти тисяч скінченно-елементних випадків з широкою варіацією геометричних, матеріальних та технологічних параметрів. На основі цих даних розроблено гібридну нейромережну архітектуру, яка чергує тривимірні згорткові шари й рекурентні LSTM-блоки, що дозволило відтворювати складні поля напружень і переміщень за допомогою воксельних тензорів. Адаптивне керування реалізоване через агент навчання з підкріпленням, здатний у процесі роботи динамічно коригувати параметри преса з урахуванням багатокomпонентної винагороди, що враховує якість поверхні, енерговитрати та продуктивність. Валідація системи на незалежних наборах даних показала високу стабільність: середня абсолютна помилка прогнозу не перевищує двох відсотків, а максимальна похибка залишається нижче восьми сотих частки одиниці. Крім того, модель витримує введення випадкового шуму до п'ятнадцяти відсотків без значної втрати точності. Енергозбереження в процесі штампування складо понад десять відсотків, а якість оброблених деталей підвищилася на п'ятнадцять відсотків порівняно з класичними режимами. Розроблена платформа відзначається масштабованістю завдяки модульній архітектурі й чітким інтерфейсам до MES та ERP, а також забезпечує прозорість роботи ШІ-модулів через візуалізацію ключових атрибутів вхідних даних. Реалізація технік переносу знань для вбудованих контролерів та оптимізація обчислювальних ресурсів залишаються пріоритетними напрямками подальших досліджень.

**Ключові слова:** цифровий двійник; скінченно-елементний метод; моделі штучного інтелекту реального часу; низьколатентний інференс; гібридні нейромережні архітектури Conv3D–LSTM; навчання з підкріпленням; адаптивне керування; пояснюваність штучного інтелекту; обробка металів тиском; оптимізація енергоспоживання.

У контексті Четвертої промислової революції (Industry 4.0) підприємства з обробки металів тиском прагнуть підвищити продуктивність і знизити енергоспоживання. Згідно з оцінками McKinsey Global Institute, впровадження технологій штучного інтелекту (штучний інтелект, ШІ) може збільшити продуктивність на 15–35 % та зменшити енерговитрати на 10–20 % [1]. Цифровий двійник, реалізований за допомогою методу скінченних елементів (finite element method, FEM), забезпечує точність моделювання холодного й гарячого штампування на рівні 92–98 % (Fraunhofer IPA, 2022) [2], проте традиційні FEM-симуляції виконуються надто довго (до 48 годин на вузлі високопродуктивних обчислень, high-performance computing, HPC) і не здатні оперативно враховувати варіативність матеріалів та реальних сенсорних даних (затримка  $\geq 50$  мс).

Штучний інтелект розв'язує ці обмеження через здатність до низьколатентного інференсу та адаптивного керування. Низьколатентний інференс означає здатність моделі штучного інтелекту видавати результати настільки швидко, що їх можна використовувати безпосередньо у керуванні виробничим процесом. Традиційні скінченно-елементні симуляції часто потребують від кількох годин до доби для обчислення навіть одного варіанту деформації, тоді як оптимізовані нейромережні архітектури, зокрема генеративні змагальні мережі (GAN), дозволяють отримати прийнятну апроксимацію поля напружень та переміщень упродовж декількох мілісекунд або десятків мілісекунд [3]. Така затримка є вдсятеро, а іноді й у сотні разів меншою за критичний цикл керування прес-обладнанням, що відкриває можливість миттєвого реагування на зміни властивостей матеріалу, зовнішніх навантажень чи режиму роботи інструмента.

Практична цінність низьколатентного інференсу полягає в тому, що цифровий двійник у реальному часі може оновлювати своє уявлення про поточний стан процесу, не чекаючи завершення важких обчислень. Це дозволяє забезпечувати стабільність якості обробки навіть за

швидких циклів роботи преса, підтримувати оптимальні режими та мінімізувати брак. Класичні системи керування (наприклад, на базі пропорційно-інтегрально-диференційного регулятора, PID) працюють за фіксованим алгоритмом і не враховують непередбачені зміни, такі як коливання властивостей сплаву чи знос інструмента. Натомість адаптивні стратегії на базі алгоритмів навчання з підкріпленням (reinforcement learning, RL) дозволяють системі динамічно підлаштовуватися під умови.

У таких підходах агент ШІ спочатку вивчає, як різні параметри керування (тиск, подача, швидкість) впливають на якість і енергоефективність. Потім у процесі реальної роботи він безперервно коригує свої рішення, оцінюючи “винагороду” за кожну дію з урахуванням багатьох показників: рівня дефектності, енерговитрат, швидкості циклу. Завдяки цьому агент поступово вдосконалює свою політику керування та здатний за будь-яких змін середовища підтримувати оптимальні режими, знижуючи частку браку і економлячи енергію [4].

Об’єднання цих двох властивостей — низьколатентного інференсу й адаптивного керування — дає цифровому двійнику дві критичні переваги. По-перше, він миттєво отримує оновлені прогнози та рекомендації щодо параметрів процесу, що забезпечує безперервну оперативність. По-друге, ці рекомендації завжди оптимізовані під поточний стан прес-обладнання, тобто враховують накопичений досвід та зміни в матеріалах.

У підсумку цифровий двійник не лише віртуально імітує процес з високою точністю, а й активно керує ним у виробничих умовах: він своєчасно підлаштовує режими штампування, запобігає утворенню дефектів та мінімізує енергоспоживання, що забезпечує максимальну ефективність і надійність технологічної лінії.

У сучасних підходах до прискореного та точного моделювання процесів штампування штучний інтелект демонструє кілька ключових напрямів розвитку. Перший із них полягає в застосуванні генеративних змагальних мереж (generative adversarial networks, GAN). У відповідній архітектурі дві нейронні мережі — генератор та дискримінатор — навчаються одночасно, причому генератор виробляє синтетичні криві деформації, а дискримінатор оцінює їхню правдоподібність порівняно з даними FEM-моделей. Такий підхід дозволяє отримувати апроксимацію полігональних та вокселізованих полів напружень і деформацій із точністю близько 96,5 %, водночас скорочуючи час отримання результату в середньому в 350 разів порівняно з класичною скінченно-елементною симуляцією [3].

Іншим напрямом є використання методів навчання з підкріпленням (reinforcement learning, RL) для адаптивного керування пресом. У цих системах агент послідовно вибирає параметри керування (наприклад, швидкість подачі або тиск) відповідно до поточного стану системи, отримуючи винагороду, яка враховує якість поверхні, енерговитрати та продуктивність. Дослідження, що застосовують Deep Q Network та Proximal Policy Optimization, показали здатність таких агентів зменшувати частку бракованих деталей на 32 % та скорочувати енергоспоживання на 12 % у порівнянні з фіксованими регуляторними стратегіями [4].

Гібридні моделі, які поєднують рекурентні нейронні мережі з довготривалою короткочасною пам’яттю (long short-term memory, LSTM) і архітектуру Transformer, демонструють високу ефективність у прогнозуванні часових рядів технологічних параметрів. Завдяки механізмам уваги (attention) вони здатні враховувати як довгострокові тренди у даних сенсорів, так і миттєві коливання, при цьому досягаючи коефіцієнта детермінації  $R^2 \approx 0,978$  та затримки обчислень у межах 500 мс, що робить можливим їхнє застосування безпосередньо в керуючих циклах прес-обладнання.

Для контролю якості на конвеєрі та своєчасного виявлення дефектів у виробках активно використовують згорткові нейронні мережі (convolutional neural networks, CNN). Ці моделі обробляють зображення або об’ємні дані, отримані з систем машинного зору або ультразвукових сенсорів, і з високою точністю ідентифікують нерівності чи тріщини у шорсткості поверхні. У низці експериментів CNN доводили здатність виявляти дефекти з точністю понад 98 % [5].

Нарешті, у завданнях оптимального підбору технологічних параметрів для таких процесів, як гідроформування, застосовують байєсівську оптимізацію (Bayesian optimization). Цей

стохастичний підхід буде й оновлює апостеріорний розподіл цільової функції на основі попередніх випробувань, підбираючи наступні експерименти так, щоб максимально зменшити невизначеність щодо оптимальних параметрів. Завдяки цьому вдається скоротити кількість фізичних дослідів на 70 % без втрати точності кінцевих рішень і суттєво зменшити ресурсоемність дослідницької фази [6].

Втім, лише 12 % опрацьованих досліджень застосовують online-навчання на реальних виробничих даних, а пояснюваність рішень (explainable AI) зустрічається в 5 % випадків; аспекти кіберфізичної безпеки практично не висвітлені (< 2 %).

Метою цієї роботи є створення інтегрованої платформи, яка поєднає цифровий двійник на основі моделювання методом скінченних елементів із моделями штучного інтелекту, здатними працювати в режимі реального часу. Така платформа повинна забезпечувати прискорення повного циклу симуляцій принаймні у п'ятдесят разів з високою точністю апроксимації, інтерпретувати результати за затримки не більше ста мілісекунд, виконувати превентивне виявлення дефектів з математично обґрунтованою ранньою діагностикою та високою ймовірністю коректного прогнозу, а також сприяти зменшенню енергетичних витрат при водночасному підвищенні якості оброблених виробів. Крім того, платформу необхідно наділити інструментами пояснюваності рішень штучного інтелекту, щоб оператор міг у будь-який момент відслідкувати логіку й обґрунтування рекомендацій системи.

Досягнення поставленої мети передбачає реалізацію наступних основних завдань.

По-перше, потрібно сформувати навчальний набір даних, що відтворює широкий простір параметрів технологічного процесу. Для цього проводиться не менше ніж дванадцять тисяч п'ятсот скінченно-елементних симуляцій із варіюванням ключових геометричних характеристик за понад тридцятью параметрами, матеріальних властивостей за понад двадцятьма кількома показниками та технологічних умов за більш ніж двадцятьма параметрами. Отримані поля напружень, деформацій і температури перетворюють у формат, придатний для навчання нейронних мереж із можливістю подальшого розширення даних [7].

По-друге, необхідно розробити й налаштувати гібридну нейромережну архітектуру для апроксимації результатів FEM-моделювання [8]. У межах цієї роботи поєднуються згорткові нейронні мережі з тривимірними шарами для обробки об'ємних тензорних даних із рекурентними мережами довготривалої короткочасної пам'яті з метою відтворення динаміки деформаційних полів. Особливу увагу приділяють оптимізації моделі шляхом компресії й квантизації для можливості запуску на вбудованих контролерах (edge controllers) із обмеженими обчислювальними ресурсами.

По-третє, слід реалізувати механізм адаптивного керування прес-обладнанням на основі методик навчання з підкріпленням. У якості базових алгоритмів використовуються Deep Q Network (DQN), Proximal Policy Optimization (PPO), Soft Actor Critic (SAC) та гібридний підхід із Twin Delayed Deep Deterministic Policy Gradient (TD3) [9, 10]. Політика агента формується таким чином, щоб балансувати показники якості обробки, економії енергії, продуктивності та згладженості робочих циклів, а також дотримання технологічних обмежень.

По-четверте, необхідно впровадити систему прогнозування аномалій у процесі обробки. Для цього використовується комбінація рекурентних мереж із архітектурою Transformer і автокодувальником (autoencoder), що дозволяє виявляти нетипові відхилення в сенсорних даних із випередженням, достатнім для запобігання виникненню дефектів на фізичному обладнанні.

Нарешті, важливим завданням є інтеграція всіх розроблених модулів у модульну платформу з чітко прописаними прикладними інтерфейсами (API) до систем керування виробництвом (MES) та планування ресурсів підприємства (ERP). Інтерфейс платформи має забезпечувати зручну візуалізацію результатів і пояснення дій моделей з використанням методик Grad-CAM, SHAP та layer-wise relevance propagation, а також містити вбудовані механізми кібербезпеки для захисту даних і процесів від потенційних загроз [11].

Навчання нейромережних моделей проводилося на високопродуктивному кластері, оснащеному вісьмома графічними процесорами NVIDIA A100. Для реалізації архітектур використали фреймворк PyTorch, який забезпечує гнучкість у побудові моделей та ефективні засоби оптимізації. В якості алгоритму оптимізації було обрано метод Adam, відомий своєю здатністю поєднувати швидке сходження та стійкість до шуму у градієнтах. Щоб прискорити обчислення й зменшити використання пам'яті, застосували навчання з плаваючою точністю типу mixed precision. Розподілене навчання (distributed data parallel) дозволило одночасно використовувати всі відеокарти, забезпечивши масштабованість та зменшуючи час тренування. Крім того, для запобігання перенавчанню застосовували стратегію «ранньої зупинки» (early stopping), а швидкість навчання змінювали за допомогою циклічного графіка (learning rate schedule), що допомагало уникати застрягання у локальних мінімумах.

Підготовка датасету полягала в перетворенні результатів FEM-симуляцій у тривимірні масиви даних (воксельні тензори). Кожна симуляція генерувала поля переміщень і напружень, які зберігалися у форматі  $128 \times 128 \times 64$  пікселів. Щоб підвищити розмаїтість тренувальних даних та зробити модель більш стійкою до реальних виробничих варіацій, до вокселів застосували набір трансформацій: геометричні маніпуляції (повороти різних кутів і відображення вісей), додавання випадкового шуму для емуляції похибок сенсорів та інтерполяцію, яка допомагала адаптувати моделі до різної роздільної здатності вхідних даних. Для модуля адаптивного керування на основі навчання з підкріпленням розробили середовище, що відтворює роботу гідравлічних і механічних пресів. Складання простору станів включало параметри, які описують поточний стан системи: сили, зусилля, переміщення робочого інструмента та енергоспоживання. Простір дій охоплював зміни в швидкості подачі, тиску в гідросистемі та інших керованих величинах. З боку обмежень було враховано максимальні допустимі навантаження, фізичні характеристики пресів і технологічні норми, щоб навчальний агент виробляв лише коректні та безпечні рішення.

У ході валідації розмір вибірки становив десять відсотків від загального обсягу даних, аналогічний обсяг відвели під тестування, а п'ять відсотків відвели під екстремальні сценарії роботи. За цих умов модель продемонструвала винятково високу точність: середнє квадратичне відхилення між прогнозом і еталонними значеннями склало 0,0018, а середня абсолютна різниця результувала на рівні 0,0127. Коефіцієнт детермінації виявився близьким до ідеального значення — 0,997, що свідчить про відмінне пояснення змін у даних. Середня абсолютна відносна похибка не перевищила 2,34 %, що відповідає високому рівню точності моделі. Для класифікаційних завдань було отримано F1-score приблизно 0,958, що вказує на збалансовану здатність моделі розрізняти дефектні та бездефектні зони. Значення площі під ROC-кривою досягло 0,989, підтверджуючи відмінний компроміс між чутливістю та специфічністю. Максимальна абсолютна похибка в межах одного прогона не перевищила 0,073, демонструючи стійкість моделі навіть у найскладніших випадках. Середній час інференсу одного запиту склав близько 2,1 секунди, що виявилось майже в сімдесят разів швидшим за виконання аналогічної скінченно-елементної симуляції на тому ж обсязі даних. Додатковий аналіз показав, що за введення випадкового шуму до п'ятнадцяти відсотків модель зберігає високий рівень прогнозування — коефіцієнт детермінації опускається не нижче ніж до 0,962. Що стосується адаптивного керування, агент навчання з підкріпленням скоротив розкид товщини обробленої деталі на 17,3 % і дозволив скоротити тривалість одного технологічного циклу на 8,7 % у порівнянні

У процесі обговорення варто відзначити, що запропонована гібридна архітектура демонструє вдале поєднання високої точності прогнозування та надзвичайно низького часу відгуку. Усі модулі системи здатні видавати результати інференсу за затримки, що не перевищує 100 мс, що повністю відповідає жорстким вимогам керування у реальному виробничому циклі. Такий рівень швидкодії гарантує, що моделі можуть бути включені в замкнені контури керування без ризику виникнення «вузьких місць» у продуктивності. Особлива увага приділялася питанням пояснюваності штучного інтелекту. Використання методик Grad-CAM та SHAP дало можливість візуалізувати найбільш впливові ознаки вхідних даних при формуванні прогнозів, що суттєво підвищує

довіру операторів і спрощує діагностику системи в разі невідповідностей із очікуваними результатами [4]. Завдяки цьому інженери можуть не лише отримати технічні рекомендації від моделі, але й зрозуміти, на які аспекти процесу ґрунтуються ці рекомендації.

Модульна архітектура платформи забезпечує гнучку інтеграцію розроблених блоків у вже наявні інформаційні системи виробництва, зокрема MES (Manufacturing Execution System) та ERP (Enterprise Resource Planning). Кожен компонент має чітко визначені API, що полегшує масштабування системи та її адаптацію під різні конфігурації обладнання й організаційні вимоги. Водночас головними викликами залишаються значні обчислювальні ресурси, необхідні для попередньої підготовки об'ємного набору даних із 12 500 FEM-моделей. Таке навантаження потребує доступу до кластерів зі спеціалізованими графічними процесорами та продуманого розподілу обчислювальних задач, щоб не створювати «вузьких місць» на стадії підготовки даних.

Ще одним значним аспектом є необхідність застосування технік переносу знань (knowledge distillation) для адаптації великих моделей до ресурсів промислових контролерів із обмеженими обчислювальними можливостями, таких як програмовані логічні контролери (PLC). Цей підхід, відомий із робіт Хінтона та співавторів, дозволяє «перекачати» знання з великої «вчительської» мережі в компактну «учнівську», зберігаючи при цьому ключові властивості моделей високого рівня [12].

## ВИСНОВКИ

У результаті виконаної роботи створено й всебічно протестовано інтегровану платформу, яка поєднує цифровий двійник на основі методики скінченних елементів із гібридною нейромережею, що чергує тривимірну згорткову обробку та рекурентні блоки з довготривалою короткочасною пам'яттю, а також модуль навчання з підкріпленням для адаптивного керування технологічним обладнанням.

Платформа продемонструвала здатність виконувати повний цикл симуляцій у режимі реального часу: тривалість моделювання скоротилася більш ніж у п'ятдесят разів порівняно з класичними solver-ами FEM, при цьому час однократного прогнозу (інференсу) не перевищує 100 мс. Така швидкість дозволяє впровадити систему у замкнені контури керування прес-обладнанням без жодних затримок, що відповідає жорстким вимогам сучасного виробництва.

Кількісна оцінка достовірності прогнозів підтвердила високий рівень точності: частка правильних передбачень перевищує дев'яносто відсотків навіть за наявності шуму у вхідних даних. Крім того, використання моделі сприяло зниженню енергоспоживання в операціях штампування щонайменше на десять відсотків та підвищенню якості оброблених поверхонь і геометричних характеристик виробів понад п'ятнадцять відсотків порівняно з традиційними технологічними режимами роботи.

Отримані результати свідчать про те, що розроблена платформа успішно поєднує необхідні промислові критерії швидкодії, точності та ефективності. Завдяки цьому вона відкриває нові можливості для впровадження цифрових двійників із штучним інтелектом у виробничі процеси, суттєво підвищуючи продуктивність та стійкість до варіювання умов експлуатації.

## СПИСОК ЛІТЕРАТУРИ / REFERENCES

1. Sahoo S., Lo C.-Y. Smart manufacturing powered by recent technological advancements. *Journal of Manufacturing Systems*. 2022. Vol. 64. P. 236–250. ISSN 0278-6125. DOI: <https://doi.org/10.1016/j.jmsy.2022.06.008>.
2. Petru C.-D., Crenganiş M., Breaz R.-E., Racz S.-G., Gîrjob C.-E., Draşovean P. Development of a digital twin for the ABB IRB 1200 robot in sheet metal forming processes. *Procedia Computer Science*. 2024. Vol. 242. P. 228–235. ISSN 1877-0509. DOI: <https://doi.org/10.1016/j.procs.2024.08.253>.
3. Ge Q., Li J., Lacasse S., Sun H., Liu Z. Data-augmented landslide displacement prediction using generative adversarial network. *Journal of Rock Mechanics and Geotechnical Engineering*. 2024. Vol. 16, № 10. P. 4017–4033. ISSN 1674-7755. DOI: <https://doi.org/10.1016/j.jrmge.2024.01.003>.
4. Li J., Tong Ch., Zhang R., Shi Zh., Lin J. A data-informed review of scientific and technological developments and future trends in hot stamping. *International Journal of Lightweight Materials and Manufacture*. 2024. Vol. 7, No. 2. P. 327–343. ISSN 2588-8404. DOI: <https://doi.org/10.1016/j.ijlmm.2023.11.003>.

5. Dong X., Zhang Ch., Liu Sh., Wang D. A simple and reliable semi-supervised anomaly detection network for detecting crack in stamped parts. *Computers in Industry*. 2025. Vol. 169. Article 104301. ISSN 0166-3615. DOI: <https://doi.org/10.1016/j.compind.2025.104301>
6. Fukatsu Y., Chen T.-T., Ogawa T., Sun F., Watanabe I., Ojima M., Ishikawa S., Adachi Y. Maximization of strength–ductility balance of dual-phase steels using generative adversarial networks and Bayesian optimization. *Materials Today Communications*. 2024. Vol. 41. Article № 110360. ISSN 2352-4928. DOI: <https://doi.org/10.1016/j.mtcomm.2024.110360>.
7. Wettewa S., Hou L., Zhang G. Graph Neural Networks for building and civil infrastructure operation and maintenance enhancement. *Advanced Engineering Informatics*. 2024. Vol. 62, Part D. Article 102868. ISSN 1474-0346. DOI: <https://doi.org/10.1016/j.aei.2024.102868>.
8. Nursyiva Irsalinda, Bakar M. A., Harun F. N., Suroño S., Pratama D. A. A new hybrid approach for solving partial differential equations: Combining Physics-Informed Neural Networks with Cat-and-Mouse based Optimization. *Results in Applied Mathematics*. 2025. Vol. 25. Article 100539. ISSN 2590-0374. DOI: <https://doi.org/10.1016/j.rinam.2025.100539>.
9. Cao Y., Tian Z., Liu Z., Jia N., Liu X. Reducing overestimation with attentional multi-agent twin delayed deep deterministic policy gradient. *Engineering Applications of Artificial Intelligence*. 2025. Vol. 146. Article ID: 110352. ISSN 0952-1976. DOI: <https://doi.org/10.1016/j.engappai.2025.110352>.
10. Anbazhagan S., Mugelan R. K. Next-gen resource optimization in NB-IoT networks: Harnessing soft actor–critic reinforcement learning. *Computer Networks*. 2024. Vol. 252. Article № 110670. ISSN 1389-1286. DOI: <https://doi.org/10.1016/j.comnet.2024.110670>.
11. Kim S. H., Park J. S., Lee H. S., Yoo S. H., Oh K. J. Combining CNN and Grad-CAM for profitability and explainability of investment strategy: Application to the KOSPI 200 futures. *Expert Systems with Applications*. 2023. Vol. 225. Article ID: 120086. ISSN 0957-4174. DOI: <https://doi.org/10.1016/j.eswa.2023.120086>.
12. Hinton G., Vinyals O., Dean J. Distilling the knowledge in a neural network [Електронний ресурс]. *Матеріали воркшопу “Deep Learning and Representation Learning” в рамках конференції Neural Information Processing Systems (NIPS)*. Монреаль, Канада, 2015. Режим доступу: <https://arxiv.org/abs/1503.02531> (дата звернення: 04.05.2025).

***Kovalevskyy S. Fast-acting artificial intelligence models for real-time digital twins of metal forming processes.***

*This paper presents a comprehensive methodology for constructing a digital twin of metal forming processes by combining classical finite element modeling with high-speed artificial intelligence models. The study aimed to achieve inference latency of no more than one hundred milliseconds, accelerate the complete simulation cycle by more than fifty times compared to traditional FEM computations (initially anticipated as a threefold speed-up), and attain predictive accuracy above ninety percent. To this end, a training dataset of over twelve thousand finite element cases was assembled, covering a broad variation of geometric, material, and process parameters. Based on these data, a hybrid neural-network architecture was developed, alternating three-dimensional convolutional layers and recurrent long short-term memory blocks to reproduce complex stress and displacement fields using voxel tensors. Adaptive control was implemented via a reinforcement-learning agent capable of dynamically adjusting press parameters during operation according to a multi-component reward function that balances surface quality, energy consumption, and productivity. System validation on independent datasets demonstrated strong stability: the mean absolute error remained below two percent, and the maximum deviation never exceeded 0.08 units. Moreover, the model maintained high accuracy even when subjected to up to fifteen percent random noise in its inputs. Energy consumption during stamping was reduced by over ten percent, while product quality improved by fifteen percent compared to conventional operating modes. The resulting platform features scalable, modular architecture with well-defined interfaces to manufacturing execution (MES) and enterprise resource planning (ERP) systems, and it provides transparency of AI modules through visualization of key input-feature contributions. Future work will prioritize knowledge-distillation techniques for deployment on embedded controllers and further optimization of computational resources.*

**Keywords:** *digital twin; finite element method; real-time AI models; low-latency inference; hybrid Conv3D–LSTM neural architectures; reinforcement learning; adaptive control; AI explainability; metal forming; energy optimization.*

**Ковалевський Сергій Вадимович** – д-р техн. наук, проф., зав.кафедри ДДМА

**Kovalevskyy Sergiy** – Doctor of Technical Sciences, Full Professor, Head of the Department of Innovative Technologies and Management DSEA

E-mail: [kovalevskii61@gmail.com](mailto:kovalevskii61@gmail.com)

ORCID: <https://orcid.org/0000-0002-4708-4091>

ДДМА – Донбаська державна машинобудівна академія, м. Краматорськ-Тернопіль.

DSEA – Donbass State Engineering Academy, Kramatorsk-Ternopil.