УДК 004.891.2

ДВУХСТУПЕНЧАТАЯ АРХИТЕКТУРА СИСТЕМЫ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ФИНАНСОВЫХ ДОКУМЕНТОВ НА ОСНОВЕ CNN И БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЕЙ

© 2026 Д.Д. Турыгина, А.Н. Жданова

Самарский национальный исследовательский университет имени академика С.П. Королёва, г. Самара, Россия

Статья поступила в редакцию 27.04.2026

В условиях цифровизации экономики и роста объёмов финансовой документации возрастает актуальность автоматизации анализа исполнения финансового плана предприятий малого и среднего бизнеса (МСП). Традиционная ручная обработка разнородных документов: договоров, актов выполненных работ, счетов-фактур и расчётных листов в форматах PDF, DOCX и растровых изображений – требует значительных временных затрат и сопровождается высокой вероятностью ошибок, обусловленных человеческим фактором. В работе предложена двухступенчатая архитектура системы интеллектуального анализа финансовых документов, объединяющая свёрточную нейронную сеть (CNN) на базе MobileNetV2 для извлечения структурированных реквизитов и большую языковую модель (LLM) для выполнения план-факт анализа и формирования интерпретируемых аналитических рекомендаций. Первая ступень выполняет извлечение и нормализацию структурированных реквизитов на основе анализа пространственной структуры документа, классификации его типа, распознавания текста средствами Tesseract OCR и последующей обработки полученных данных. Вторая ступень выполняет сопоставление извлечённых фактов с табличными данными финансового плана, вычисляет отклонения по статьям и периодам, формирует интегральную оценку качества исполнения плана и генерирует текстовые пояснения и управленческие рекомендации. Отличительной особенностью предлагаемой архитектуры является использование структурированных фактов, сформированных на первой ступени, в качестве документального основания выводов второй ступени. Таким образом проверяются аналитические заключения и выявляются расхождения между финансовым планом и первичным документооборотом. Экспериментальная апробация проведена на выборке из 100 реальных финансовых документов предприятия ООО «СПЕЦПРОМСТРОЙ» за период 2023-2025 гг. и 400 синтетически сгенерированных документов. Точность распознавания текста составила 92%, точность анализа разметки документа – 94%, а точность нормализации извлечённых полей – 85%. Применение разработанной системы сокращает время обработки на 42% (с 10 до 5,8 ч при обработке 30 документов) и снижает количество ошибок на 24% по сравнению с ручной обработкой аналогичного объёма данных. Полученные результаты доказывают эффективность предложенного архитектурного решения для автоматизации анализа финансовых документов предприятий МСП в условиях ограниченных вычислительных и кадровых ресурсов.

Ключевые слова: свёрточные нейронные сети, большие языковые модели, финансовый анализ, план-факт анализ, извлечение информации из документов, автоматизация, малый и средний бизнес.

DOI: 10.37313/1990-5378-2026-28-4-202-211
EDN: JJSAMS

ВВЕДЕНИЕ

Анализ исполнения финансового плана является одной из основных задач управления предприятием, заключающейся в сопоставлении плановых показателей бюджета с фактическими данными о доходах и расходах, выявлении причин отклонений и выработке корректирующих управленческих решений [1]. На предприятиях малого и среднего бизнеса (МСП) организация данного процесса имеет ряд особенностей: анализ выполняется небольшим числом сотрудников (как правило, бухгалтером или финансовым менеджером) без выделенной аналитической службы; специализированные корпоративные системы бюджетирования зачастую недоступны из-за высокой стоимости внедрения и сопровождения; ограничены вычислительные ресурсы [2].

Исходными данными для анализа служит первичный документооборот: договоры, счета-фактуры, акты выполненных работ, расчётные листы, представленные в различных форматах (PDF, DOCX, Excel, сканированные изображения) и не имеющие унифицированной структуры [3]. Сопоставление плановых и фактических показателей в ручном режиме требует последовательного решения двух взаимосвязанных задач: сначала извлечения реквизитов из неструктурированных

Турыгина Дарья Денисовна, магистрант института информатики и кибернетики. E-mail: ddturygina@mail.ru
Жданова Александра Николаевна, кандидат технических наук, доцент кафедры программных систем. E-mail: danilenko.an@ssau.ru

документов, затем проведения собственно финансового анализа. По экспертным оценкам, ручная обработка сопровождается ошибками в 10–15 % документов и занимает около 8 часов на обработку 100 документов.

Методы машинного обучения демонстрируют высокую эффективность при решении обеих указанных задач [4, 5]. Свёрточные нейронные сети (CNN) успешно применяются для анализа визуальной структуры документов и классификации их типов [6-8], а большие языковые модели (LLM) – для интерпретации числовых показателей и генерации аналитических текстов [9].

Однако существующие решения, как правило, автоматизируют лишь отдельные этапы обработки: либо извлечение информации из документов, либо аналитическую обработку уже структурированных данных, не обеспечивая сквозного процесса «от документа до рекомендации». Кроме того, большинство систем интеллектуальной обработки документов разрабатывается для крупных организаций и не учитывает специфику предприятий МСП, связанную с ограниченными вычислительными ресурсами, отсутствием стандартизации документооборота и небольшими объёмами обучающих данных.

Двухэтапная организация процесса анализа, включающая первоначальную сверку документов и последующую интерпретацию выявленных отклонений, обусловливает построение двухступенчатой архитектуры системы, в которой первая ступень воспроизводит функции оператора по извлечению реквизитов, а вторая — функции аналитика по объяснению отклонений. Целью настоящей работы является разработка и экспериментальная оценка эффективности двухступенчатой архитектуры на основе свёрточных нейронных сетей и больших языковых моделей для автоматизации анализа исполнения финансового плана предприятий малого и среднего бизнеса.

Для достижения поставленной цели решаются следующие задачи:

1. Разработка архитектуры двухступенчатого конвейера обработки, обеспечивающего модульность и раздельную оценку качества каждой ступени.
2. Реализация CNN-модуля извлечения структурированных реквизитов из неструктурированных финансовых документов.
3. Реализация LLM-модуля план-факт анализа с генерацией интерпретируемых аналитических рекомендаций.
4. Экспериментальная оценка эффективности системы на реальных данных предприятия МСП.
5. Количественное сравнение двухступенчатой архитектуры с одноступенчатой архитектурой обработки.

Научная новизна работы заключается в способе интеграции свёрточной нейронной сети и большой языковой модели, при котором структурированные факты, извлечённые первой ступенью, передаются второй ступени в качестве документального основания формируемых выводов. Предложенный подход обеспечивает ряд преимуществ по сравнению с одноступенчатой обработкой: во-первых, каждое аналитическое утверждение связывается с подтверждающим документом, что снижает вероятность формирования необоснованных выводов языковой моделью; во-вторых, появляется возможность выявления расхождений между финансовым планом и документооборотом, включая бюджетные статьи без подтверждающих документов и договоры, не отражённые в плане; в-третьих, качество работы каждой ступени оценивается независимо, что упрощает локализацию и исправление ошибок.

ОБЗОР СУЩЕСТВУЮЩИХ ПОДХОДОВ

Задача извлечения структурированных данных из неструктурированных документов активно развивается в последние годы [10]. Современные подходы объединяют методы оптического распознавания символов [11] и глубокого обучения, учитывающие пространственную структуру (разметку) документа.

Наиболее близкими к настоящей работе являются архитектуры, объединяющие визуальный и текстовый анализ документа. Семейство моделей LayoutLM [12, 13] выполняет совместное предобучение на тексте, координатах и изображении страницы и стало одним из наиболее распространённых решений для понимания структуры деловых документов. Модель Donut [14] реализует сквозное извлечение структурированных данных непосредственно из изображения документа без отдельного этапа распознавания символов. Модель DocLLM [15], ориентированная на обработку корпоративных и финансовых документов, дополняет большую языковую модель информацией о пространственной структуре страницы. Общим ограничением перечисленных моделей являются высокая вычислительная сложность, значительное число параметров и необходимость использования графических ускорителей, что затрудняет их практическое применение на предприятиях малого и среднего бизнеса.

В финансовой сфере большие языковые модели используются для анализа тональности финансовой отчётности (FinBERT [16]), решения задач числового анализа финансовых таблиц (FinQA [17]) и комплексной финансовой аналитики (BloombergGPT [9]). Однако перечисленные подходы ориентированы преимущественно на обработку уже структурированных или текстовых данных и не решают задачу извлечения реквизитов из первичных финансовых документов. Работы, объединяющие извлечение реквизитов и план-факт анализ в рамках единого сквозного процесса применительно к документообороту предприятий МСП, авторам неизвестны. Предлагаемая архитектура направлена на устранение данного ограничения.

Для реализации первой ступени выбрана компактная архитектура MobileNetV2 [18], обеспечивающая рациональное соотношение точности классификации и вычислительной эффективности благодаря использованию инвертированных остаточных блоков. По сравнению с исходной архитектурой MobileNet [19] она требует меньшего объёма вычислений при сопоставимой точности, что особенно важно при развёртывании системы в условиях ограниченных вычислительных ресурсов предприятий МСП.

Практическая эффективность применения больших языковых моделей в финансовых процессах подтверждается отраслевыми исследованиями [20]. Во второй ступени используется русскоязычная модель семейства GPT-2 (RuGPT-2), выбор которой обусловлен возможностью локального развёртывания и последующего дообучения без применения специализированного вычислительного оборудования.

Модульные архитектуры, в которых извлечение информации и её последующий анализ реализуются раздельными компонентами, получили широкое распространение в системах интеллектуальной обработки данных [10]. Их основными преимуществами являются возможность независимой оценки качества и модернизации отдельных модулей, а также повышение интерпретируемости результатов за счёт явного разделения этапов обработки.

АРХИТЕКТУРА СИСТЕМЫ

Общая архитектура двухступенчатого конвейера обработки

Предлагаемая система реализована в виде двух последовательно связанных ступеней (рис. 1): извлечения структурированных фактов из финансовых документов и план-факт анализа с формированием интерпретируемых аналитических выводов. Отличительной особенностью архитектуры является передача структурированных фактов, сформированных на первой ступени, второй ступени в качестве документального основания аналитических выводов, а также реализация механизмов контроля дубликатов и привязки извлечённых фактов к статьям финансового плана. Тем самым повышается проверяемость результатов анализа и обеспечивается выявление расхождений между финансовым планом и первичным документооборотом. Модульная организация архитектуры обеспечивает раздельную оценку точности извлечения реквизитов и качества аналитики отклонений.

Система включает следующие функциональные компоненты:

- модуль аутентификации и разграничения доступа (токены JWT, ролевая модель «администратор/сотрудник», журнал безопасности);
- CNN-модуль извлечения реквизитов с возможностью дообучения на новых типах документов [7];
- LLM-модуль формирования аналитических рекомендаций [9, 15].

Ступень 1 – Извлечение реквизитов на основе CNN и распознавания текста

Первая ступень предназначена для преобразования неструктурированного документа в набор структурированных полей, необходимых для идентификации финансового события и его привязки к статьям бюджета. На вход подаётся файл документа (договор, акт, счёт-фактура), на выходе формируется карточка документа в формате JSON, содержащая нормализованные значения реквизитов и ссылку на исходный файл.

Из каждого документа извлекаются следующие реквизиты:

- контрагент (наименование организации или стороны договора);
- даты (начало и окончание действия, дата заключения);
- сумма обязательства;
- тип договора (доходы или расходы) для отнесения к соответствующему финансовому потоку.

CNN применяется для анализа визуального представления документа на уровне страницы и выделения областей, содержащих основные реквизиты.

Алгоритм обработки включает следующие этапы:

1. Конвертация документа в растровое изображение размером 224×224 пикселя.

2. Классификация и выделение основных полей с помощью CNN на основе архитектуры MobileNetV2 (TensorFlow/Keras): классификация типа документа (доход/расход), регрессия для извлечения числовых полей (сумма, даты) [18].

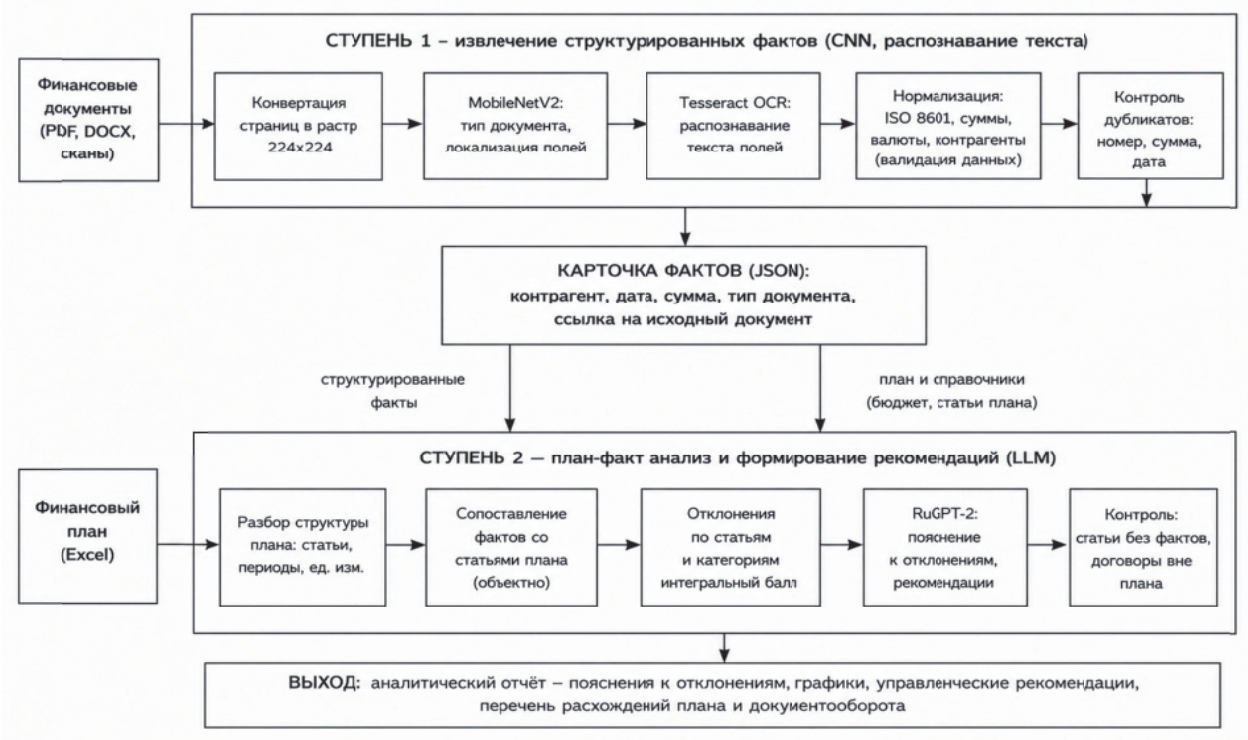

Рис. 1 – Архитектура двухступенчатого конвейера обработки финансовых документов

3. Распознавание текста посредством Tesseract OCR [11] с последующей нормализацией: даты приводятся к формату ISO 8601, числовые значения – к формату с плавающей точкой, наименования контрагентов сопоставляются со справочником методом нечёткого сравнения строк.

4. Проверка дубликатов по комбинации «номер, сумма, дата».

5. Формирование результирующей записи – карточки документа в формате JSON, передаваемой второй ступени.

Ступень 2 – План-факт анализ и формирование рекомендаций на основе LLM

Вторая ступень ориентирована на обработку табличного финансового плана и фактических значений, а также на формирование аналитического вывода. На данном этапе система выполняет:

- разбор структуры Excel-таблицы (статьи бюджета, периоды, единицы измерения);
- нормализацию наименований статей;
- сопоставление «план–факт» по идентификатору статьи и периоду;
- вычисление отклонений по каждой паре «статья–период».

LLM-модуль реализован на базе русскоязычной модели семейства GPT-2 (RuGPT-2) и формирует текстовые пояснения на основе рассчитанных отклонений и контекста, полученного из извлечённых фактов первой ступени (контрагент, сумма, тип обязательства, временной интервал). Использование фактов из документов обеспечивает:

- сопоставление значений по сумме, периоду и типу с документальными основаниями;
- выявление бюджетных статей без подтверждающих документов и договоров, не отражённых в плане или факте;
- группировку отклонений по контрагентам и типам обязательств для формирования приоритетов проверки.

Дополнительно реализована адаптивная группировка по периоду: при горизонте свыше 90 дней – помесячная агрегация, от 15 до 90 дней – понедельная, менее 14 дней – подневная.

Показатели качества исполнения финансового плана

Для количественной оценки исполнения финансового плана используются классические показатели финансового анализа, включая коэффициенты ликвидности, рентабельности и покрытия долга, рассчитываемые по стандартным методикам [21]. В отличие от традиционного подхода предлагаемая методика дополнена показателем доли аномальных операций, единой процедурой нормировки показателей и интегральным баллом качества исполнения плана, используемым при формировании аналитических рекомендаций.

Доля аномальных операций:

$$A = \frac{|\{x_i : |x_i - \mu| > 2\sigma\}|}{N},$$

где μ и σ – среднее и стандартное отклонение сумм по статьям.

Каждый показатель нормализуется в оценку $s_k \in [0,1]$, где 0 соответствует критическому состоянию, 1 – норме. Интегральный балл качества вычисляется как:

$$Q = \sum_k w_k \cdot s_k, \sum_k w_k = 1.$$

Рекомендации формируются для статей, корректировка которых повышает интегральный балл Q более чем на пороговое значение δ.

Отклонение по доходной части:

$$\Delta_{revenue} = \frac{\text{Факт}_{revenue} - \text{План}_{revenue}}{\text{План}_{revenue}} \times 100\%.$$

Положительное значение свидетельствует о перевыполнении плана.

Отклонение по расходной части:

$$\Delta_{expense} = \frac{\text{План}_{expense} - \text{Факт}_{expense}}{\text{План}_{expense}} \times 100\%.$$

Положительное значение свидетельствует об экономии средств.

Привязка фактов к плановым записям осуществляется по атрибуту linked_planned_contract, обеспечивая обработку исключительно документально подтверждённых статей.

ЭКСПЕРИМЕНТАЛЬНАЯ ОЦЕНКА

Условия эксперимента

Экспериментальная апробация проведена на базе ООО «СПЕЦПРОМСТРОЙ». Для оценки использована выборка из 100 реальных и 400 синтетически сгенерированных финансовых документов за период 2023-2025 гг., включающий договоры, акты выполненных работ, счета-фактуры в форматах PDF, DOCX и сканированные изображения.

Результаты извлечения реквизитов (ступень 1)

Точность работы первой ступени оценивалась по компонентам конвейера (таблица 1).

Таблица 1 – Точность извлечения реквизитов на выборке из 500 документов

Этап обработки	Точность	Комментарий
Распознавание текста (OCR)	92%	500 документов
CNN: классификация типа документа и элементов разметки	94%	входное изображение 224×224
Нормализация дат	98%	приведение к формату ISO 8601
Нормализация полей в целом	85%	даты, суммы, контрагенты

Наибольшая точность достигнута при нормализации дат (98%), тогда как общая точность нормализации полей составила 85%, что обусловлено разнообразием форматов записи сумм и наименований контрагентов. Общая точность распознавания текста составила 92%, а классификации типа документа и элементов разметки – 94%.

Динамика обучения CNN-модуля (MobileNetV2) представлена на рисунке 2. График отражает зависимость функции потерь и точности на обучающей и валидационной выборках от числа эпох.

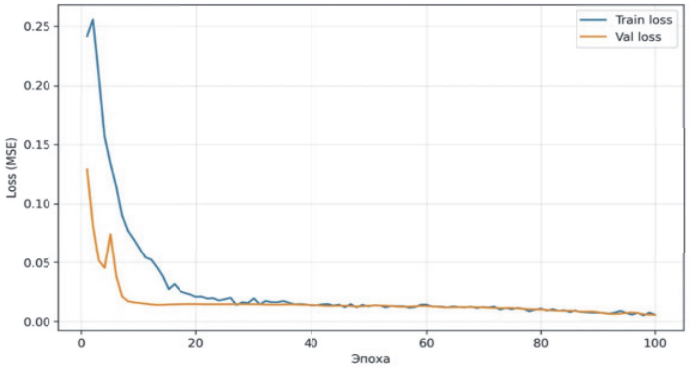

Рис. 2 – Кривые обучения CNN-модуля

Матрица ошибок классификации типов документов (доход/расход) представлена на рисунке 3.

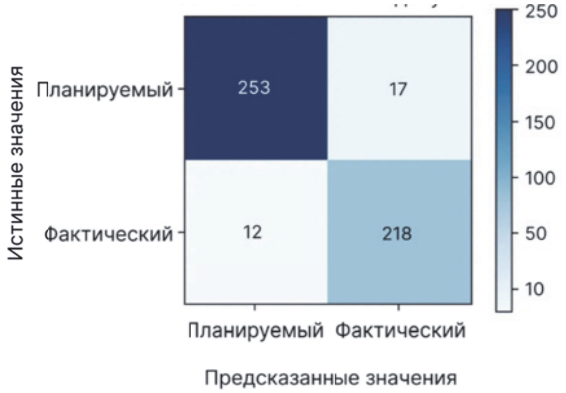

Рис. 3 – Матрица ошибок классификации типов документов

На рисунке 4 представлено сравнение качества извлечения отдельных типов реквизитов (контрагент, дата, сумма, тип договора). Для оценки качества классификации элементов разметки использовались метрики Accuracy, Precision, Recall и F1. Предложенная модель на базе MobileNetV2 достигла общей точности классификации (Accuracy) 94%, при средних по классам значениях Precision = 0,91, Recall = 0,88 и F1 = 0,89 (классы «Контент», «Таблица», «Сумма», «Текст»). Наилучший результат получен для детекции табличных блоков «Сумма» (F1 = 0,93), что особенно важно при обработке финансовых документов предприятий малого и среднего бизнеса. Полученные результаты доказывают пригодность предложенной модели для последующего распознавания текста и извлечения структурированных реквизитов.

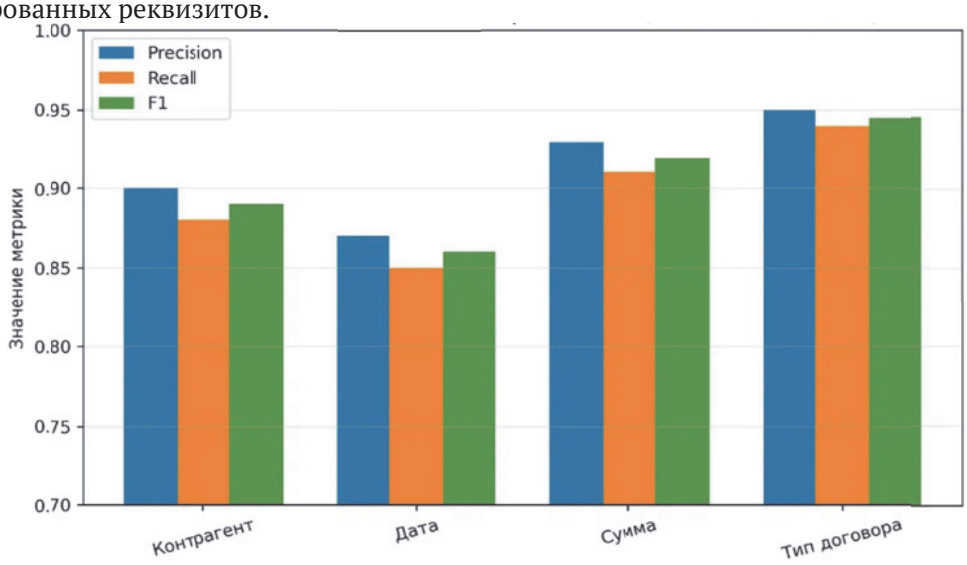

Рис. 4 – Точность извлечения по типам реквизитов (Precision, Recall, F1)

Сравнение двухступенчатой и одноступенчатой архитектур

Для количественной оценки эффективности двухступенчатой архитектуры проведено сравнение с одноступенчатой архитектурой, в которой текст, распознанный средствами Tesseract OCR, передаётся языковой модели непосредственно, без формирования промежуточного структурированного представления в виде карточек фактов. Обе архитектуры оценивались на одной и той же выборке из 500 документов, включавшем 100 реальных и 400 синтетически сгенерированных документов, по трём критериям: качеству извлечения основных реквизитов (F1-мера), доле аналитических утверждений, не имеющих документального подтверждения, и числу выявленных расхождений между финансовым планом и первичным документооборотом (таблица 2).

Использование промежуточного структурированного представления повышает качество извлечения реквизитов (F1: 0,89 вместо 0,74), существенно снижает долю аналитических утверждений без документального подтверждения и обеспечивает выявление расхождений между финансовым планом и первичным документооборотом, недоступных при одноступенчатой обработке.

Результаты план-факт анализа и практическая значимость (ступень 2)

Практическая значимость системы оценивалась сравнением автоматизированного режима с ручным по двум критериям (табл. 3):

Таблица 2 – Сравнение одноступенчатой и двухступенчатой архитектур

Критерий	Одноступенчатая архитектура	Двухступенчатая архитектура	Примечание
Точность извлечения реквизитов (F1)	0,74	0,89	500 документов
Доля утверждений без документального основания	18%	2-4%	экспертная оценка выборки пояснений
Выявленные расхождения «план–документооборот»	1-3	7	за период 2023–2025 гг.

- время обработки: обработка 30 документов сократилась с 10 часов (ручной режим) до 4,6 часов (автоматизированный режим), что составляет снижение на 42%;

- качество анализа: количество ошибок и неточностей сократилось на 24% по сравнению с ручным режимом.

Таблица 3 – Сравнение ручного и автоматизированного режимов обработки

Критерий	Ручной режим	Автоматизированный режим	Улучшение (%)
Время обработки 30 док.	10 часов	5,8 часа	42%
Количество ошибок	1 в каждом 10 док.	1 в каждом 13.2 док.	24%

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Полученные результаты подтверждают выдвинутую гипотезу исследования: двухступенчатый конвейер сокращает время сверки финансового плана на 42% (при пороговом значении 40%), а качество извлечения реквизитов (F1 = 0,89) превышает установленный порог 0,85. Общая точность классификации элементов разметки составила 94%, а распознавания текста – 92%, что подтверждает применимость выбранных архитектур к обработке финансовых документов предприятий малого и среднего бизнеса. Наименьшая точность достигнута на этапе нормализации реквизитов (85%), это связано с вариативностью форматов записи дат, денежных сумм и наименований контрагентов в документах различных организаций.

Сравнение с одноступенчатой архитектурой показало, что использование промежуточного структурированного представления повышает качество извлечения реквизитов (F1: 0,89 вместо 0,74) и обеспечивает документальную обоснованность формируемых аналитических выводов. Кроме того, предложенная архитектура способствует выявлению расхождений между финансовым планом и первичным документооборотом, которые не обнаруживаются при непосредственной передаче распознанного текста языковой модели.

Модульная организация конвейера создаёт возможность независимого развития отдельных компонентов системы. Замена MobileNetV2 на более современную архитектуру (например, EfficientNet) или обновление используемой языковой модели не требуют модификации остальных модулей, что упрощает сопровождение и последующую модернизацию системы.

ЗАКЛЮЧЕНИЕ

В работе предложена и реализована двухступенчатая архитектура системы интеллектуального анализа финансовых документов для автоматизации анализа исполнения финансового плана предприятий малого и среднего бизнеса. Архитектура объединяет свёрточную нейронную сеть для извлечения структурированных реквизитов из неструктурированных документов и большую языковую модель для выполнения план-факт анализа и формирования интерпретируемых аналитических рекомендаций.

Основные результаты работы:

1. Разработана модульная архитектура системы, обеспечивающая последовательную обработку финансовых документов и раздельную оценку качества каждого этапа обработки.

2. Разработан модуль извлечения реквизитов на основе MobileNetV2 и Tesseract OCR, обеспечивающий общую точность классификации элементов разметки 94% и точность распознавания текста 92%.

3. Разработан модуль план-факт анализа на основе большой языковой модели, реализующий расчёт финансовых показателей, формирование аналитических выводов и генерацию рекомендаций.

4. Проведено количественное сравнение двухступенчатой и одноступенчатой архитектур, показавшее повышение качества извлечения реквизитов (F1: 0,89 вместо 0,74), документальную обо-

снованность аналитических выводов и возможность выявления расхождений между финансовым планом и первичным документооборотом.

Экспериментальная апробация на выборке из 500 документов, включавшем 100 реальных и 400 синтетически сгенерированных документов, подтвердила сокращение времени обработки на 42% и снижение количества ошибок на 24%.

Полученные результаты подтверждают эффективность предложенной двухступенчатой архитектуры для автоматизации анализа финансовых документов предприятий малого и среднего бизнеса в условиях ограниченных вычислительных и кадровых ресурсов.

Направления дальнейших исследований связаны с расширением перечня обрабатываемых документов, увеличением объёма обучающей выборки для повышения качества нормализации реквизитов, а также интеграцией механизмов обратной связи, обеспечивающих адаптацию языковой модели к особенностям документооборота конкретного предприятия.

СПИСОК ЛИТЕРАТУРЫ

1. *Черепашков, А.А.* Методика оценки эффективности комплексных решений промышленной автоматизации / А.А. Черепашков, П.А. Самойлов // Известия Самарского научного центра Российской академии наук. – 2021. – Т. 23. – № 3. – С. 13-17.
2. *Барсегян, Н.В.* Направления совершенствования организации промышленного производства на основе инструментов Big Data / Н.В. Барсегян, А.И. Ван // Известия Самарского научного центра Российской академии наук. – 2025. – Т. 27. – № 6. – С. 116-123.
3. *Иванов, М.В.* Исследование структуры документооборота предприятия с помощью искусственных нейронных сетей / М.В. Иванов // Известия Самарского научного центра Российской академии наук. – 2019. – Т. 21. – № 3. – С. 41-46.
4. *Горбачева, Т.А.* Направления использования машинного обучения в финансовой индустрии / Т.А. Горбачева // Вестник Алтайской академии экономики и права. – 2025. – № 2-2. – С. 157-163;
5. *Кантаев, Н. К.* Перспективы использования методов машинного обучения для оценки кредитных рисков / Н. К. Кантаев // Вестник евразийской науки. – 2025. – Т. 17. – № s2. – URL: https://esj.today/PDF/70FAVN225.pdf (дата обращения: 24.02.2026).
6. *Свирина, А. Р.* Классификация изображений с помощью сверточных нейронных сетей / А. Р. Свирина // Молодой ученый. – 2024. – № 24 (523). – С. 96-99. – URL: https://moluch.ru/archive/523/115598 (дата обращения: 24.02.2026).
7. *Котюжанский, Л. А.* Классификация сканированных документов с использованием сверточной нейросети / Л. А. Котюжанский, Н. В. Четверкин, А. А. Протасевич [и др.] // Современные наукоемкие технологии. – 2021. – № 6 (часть 1). – С. 45-49.
8. *Сидорова, М.П.* Интеллектуальная обработка финансовых документов на основе CNN / М.П. Сидорова // Информатика и ее применения. – 2023. – Т. 17. – № 3. – С. 34-41.
9. *Wu S., Irsoy O., Lu S. et al.* BloombergGPT: A Large Language Model for Finance // arXiv preprint arXiv:2303.17564. 2023.
10. *Cui L., Xu Y., Lv T., Wei F.* Document AI: Benchmarks, Models and Applications // arXiv preprint arXiv:2111.08609. 2021.
11. *Smith R.* An Overview of the Tesseract OCR Engine // Proceedings of the 9th International Conference on Document Analysis and Recognition (ICDAR 2007). 2007. P. 629-633.
12. *Xu Y., Li M., Cui L. et al.* LayoutLM: Pre-training of Text and Layout for Document Image Understanding // Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2020. P. 1192-1200.
13. *Huang Y., Lv T., Cui L. et al.* LayoutLMv3: Pre-training for Document AI with Unified Text and Image Masking // Proceedings of the 30th ACM International Conference on Multimedia. 2022. P. 4083-4091.
14. *Kim G., Hong T., Yim M. et al*. OCR-free Document Understanding Transformer // Computer Vision – ECCV 2022. Lecture Notes in Computer Science. Vol. 13688. Springer, 2022. P. 498-517.
15. *Wang D., Raman N., Sibue M. et al.* DocLLM: A Layout-Aware Generative Language Model for Multimodal Document Understanding // arXiv preprint arXiv:2401.00908. 2024.
16. *Araci D.* FinBERT: Financial Sentiment Analysis with Pre-trained Language Models // arXiv preprint arXiv:1908.10063. 2019.
17. *Chen Z., Chen W., Smiley C. et al.* FinQA: A Dataset of Numerical Reasoning over Financial Data // Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2021. P. 3697-3711.
18. *Sandler M., Howard A., Zhu M. et al.* MobileNetV2: Inverted Residuals and Linear Bottlenecks // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2018. P. 4510-4520.
19. *Howard A.G., Zhu M., Chen B. et al.* MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications // arXiv preprint arXiv:1704.04861. 2017.
20. Исследование Альфа-Банка: считаем финансовый эффект от ИИ, ML и LLM в банке // Хабр. 2025. URL: https://habr.com/ru/companies/alfa/articles/970596/ (дата обращения: 24.03.2026).
21. *Ковалёв, В.В.* Финансовый анализ: методы и процедуры / В.В. Ковалёв. – М.: Финансы и статистика, 2002. – 560 с.

A TWO-STAGE ARCHITECTURE FOR A FINANCIAL DOCUMENT MINING SYSTEM BASED ON CNN AND LARGE LANGUAGE MODELS

© 2026 D.D. Turygina, A.N. Zhdanova

Samara National Research University named after Academician S.P. Korolyov, Samara, Russia

With the digitalization of the economy and the growth of financial documentation, the need to automate the analysis of financial plan execution in small and medium-sized enterprises (SMEs) is growing. Traditional manual processing of disparate documents—contracts, certificates of completion, invoices, and payslips in PDF, DOCX, and raster image formats—is time-consuming and is prone to human error. This paper proposes a two-stage architecture for a financial document mining system. It combines a MobileNetV2-based convolutional neural network (CNN) for extracting structured attributes and a large language model (LLM) for performing plan-fact analysis and generating interpretable analytical recommendations. The first stage extracts and normalizes structured attributes based on an analysis of the document's spatial structure, document type classification, text recognition using Tesseract OCR, and subsequent processing of the resulting data. The second stage compares the extracted facts with the financial plan's tabular data, calculates variances by item and period, generates an integrated assessment of the plan's performance quality, and generates text explanations and management recommendations. A distinctive feature of the proposed architecture is the use of structured facts generated in the first stage as a documentary basis for the conclusions of the second stage. This verifies analytical conclusions and identifies discrepancies between the financial plan and the primary document flow. The pilot was conducted on a sample of 100 real financial documents of SPETSPROMSTROY LLC for the period 2023-2025 and 400 synthetically generated documents. The text recognition accuracy was 92%, the document markup analysis accuracy was 94%, and the normalization accuracy of extracted fields was 85%. The developed system reduced processing time by 42% (from 10 to 5.8 hours when processing 30 documents) and reduced errors by 24% compared to manual processing of a similar volume of data. The obtained results demonstrate the effectiveness of the proposed architectural solution for automating the analysis of financial documents of SMEs with limited computing and human resources.

Keywords: convolutional neural networks, large-scale language models, financial analysis, plan-fact analysis, information extraction from documents, automation, small and medium businesses.

DOI: 10.37313/1990-5378-2026-28-4-202-211
EDN: JJSAMS

REFERENCES

1. *Cherepashkov, A.A.* Metodika ocenki effektivnosti kompleksnyh reshenij promyshlennoj avtomatizacii / A.A. Cherepashkov, P.A. Samojlov // Izvestiya Samarskogo nauchnogo centra Rossijskoj akademii nauk. – 2021. – T. 23. – № 3. – S. 13-17.
2. *Barsegyan, N.V.* Napravleniya sovershenstvovaniya organizacii promyshlennogo proizvodstva na osnove instrumentov Big Data / N.V. Barsegyan, A.I. Van // Izvestiya Samarskogo nauchnogo centra Rossijskoj akademii nauk. – 2025. – T. 27. – № 6. – S. 116-123.
3. *Ivanov, M.V.* Issledovanie struktury dokumentooborota predpriyatiya s pomoshch'yu iskusstvennyh nejronnyh setej / M.V. Ivanov // Izvestiya Samarskogo nauchnogo centra Rossijskoj akademii nauk. – 2019. – T. 21. – № 3. – S. 41-46.
4. *Gorbacheva, T.A.* Napravleniya ispol'zovaniya mashinnogo obucheniya v finansovoj industrii / T.A. Gorbacheva // Vestnik Altajskoj akademii ekonomiki i prava. – 2025. – № 2-2. – S. 157-163;
5. *Kantaev, N. K.* Perspektivy ispol'zovaniya metodov mashinnogo obucheniya dlya ocenki kreditnyh riskov / N. K. Kantaev // Vestnik evrazijskoj nauki. – 2025. – T. 17. – № s2. – URL: https://esj.today/PDF/70FAVN225.pdf (data obrashcheniya: 24.02.2026).
6. *Svirina, A. R.* Klassifikaciya izobrazhenij s pomoshch'yu svertochnyh nejronnyh setej / A. R. Svirina // Molodoj uchenyj. – 2024. – № 24 (523). – S. 96-99. – URL: https://moluch.ru/archive/523/115598 (data obrashcheniya: 24.02.2026).
7. *Kotyuzhanskij, L. A.* Klassifikaciya skanirovannyh dokumentov s ispol'zovaniem svertochnoj nejroseti / L. A. Kotyuzhanskij, N. V. Chetverkin, A. A. Protasevich [i dr.] // Sovremennye naukoemkie tekhnologii. – 2021. – № 6 (chast' 1). – S. 45-49.
8. *Sidorova, M.P.* Intellektual'naya obrabotka finansovyh dokumentov na osnove CNN / M.P. Sidorova // Informatika i ee primeneniya. – 2023. – T. 17. – № 3. – S. 34-41.
9. *Wu S., Irsoy O., Lu S. et al.* BloombergGPT: A Large Language Model for Finance // arXiv preprint arXiv:2303.17564. 2023.
10. *Cui L., Xu Y., Lv T., Wei F.* Document AI: Benchmarks, Models and Applications // arXiv preprint arXiv:2111.08609. 2021.
11. *Smith R.* An Overview of the Tesseract OCR Engine // Proceedings of the 9th International Conference on Document Analysis and Recognition (ICDAR 2007). 2007. P. 629-633.
12. *Xu Y., Li M., Cui L. et al.* LayoutLM: Pre-training of Text and Layout for Document Image Understanding // Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2020. P. 1192-1200.
13. *Huang Y., Lv T., Cui L. et al.* LayoutLMv3: Pre-training for Document AI with Unified Text and Image Masking // Proceedings of the 30th ACM International Conference on Multimedia. 2022. P. 4083-4091.

14. *Kim G., Hong T., Yim M. et al.* OCR-free Document Understanding Transformer // Computer Vision – ECCV 2022. Lecture Notes in Computer Science. Vol. 13688. Springer, 2022. P. 498-517.
15. *Wang D., Raman N., Sibue M. et al.* DocLLM: A Layout-Aware Generative Language Model for Multimodal Document Understanding // arXiv preprint arXiv:2401.00908. 2024.
16. *Araci D.* FinBERT: Financial Sentiment Analysis with Pre-trained Language Models // arXiv preprint arXiv:1908.10063. 2019.
17. *Chen Z., Chen W., Smiley C. et al.* FinQA: A Dataset of Numerical Reasoning over Financial Data // Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2021. P. 3697-3711.
18. *Sandler M., Howard A., Zhu M. et al.* MobileNetV2: Inverted Residuals and Linear Bottlenecks // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2018. P. 4510-4520.
19. *Howard A.G., Zhu M., Chen B. et al.* MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications // arXiv preprint arXiv:1704.04861. 2017.
20. Issledovanie Al'fa-Banka: schitaem finansovyj effekt ot II, ML i LLM v banke // Habr. 2025. URL: https://habr.com/ru/companies/alfa/articles/970596/ (data obrashcheniya: 24.03.2026).
21. *Kovalyov, V.V.* Finansovyj analiz: metody i procedury / V.V. Kovalyov. – M.: Finansy i statistika, 2002. – 560 s.

Daria Turygina, Master's Student at the Institute of Informatics and Cybernetics. E-mail: ddturygina@mail.ru
Aleksandra Zhdanova, PhD in Engineering, Associate Professor at the Department of Software Systems.
E-mail: danilenko.an@ssau.ru