

A Graph-Augmented Liquid Neural Network for Extracting Food Hazards and Disease Outbreaks

Diya Saha^{1,2}, Sudeshna Jana^{1,2}, Manjira Sinha¹, Ishan Verma¹, Vaishali Aggarwal¹,
Tirthankar Dasgupta¹

¹TCS Research and Innovation Lab, ²Indian Institute of Technology, Kharagpur

Correspondence: {diya.saha, sudeshna.jana, sinha.manjira, ishan.verma, aggarwal.vaishali, dasgupta.tirthankar}@tcs.com

Abstract

The increasing frequency of foodborne illnesses, safety hazards, and disease outbreaks in the food supply chain demands urgent attention to protect public health. These incidents, ranging from contamination to intentional adulteration of food and feed, pose serious risks to consumers, leading to poisoning, and disease outbreaks that lead to product recalls. Identifying and tracking the sources and pathways of contamination is essential for timely intervention and prevention. This paper explores the use of social media and regulatory news reports to detect food safety issues and disease outbreaks. We present an automated approach leveraging a multi-task sequence labeling and sequence classification model that uses a liquid time-constant neural network augmented with a graph convolution network to extract and analyze relevant information from social media posts and official reports. Our methodology includes the creation of annotated datasets of social media content and regulatory documents, enabling the model to identify foodborne infections and safety hazards in real-time. Preliminary results demonstrate that our model outperforms baseline models, including advanced large language models like LLaMA-3 and Mistral-7B, in terms of accuracy and efficiency. The integration of liquid neural networks significantly reduces computational and memory requirements, achieving superior performance with just $1.2 \times e^6$ bytes of memory, compared to the 20.3 GB of GPU memory needed by traditional transformer-based models. This approach offers a promising solution for leveraging social media data in monitoring and mitigating food safety risks and public health threats.

1 Introduction

The escalating number of food safety concerns remains a source of significant apprehension (Amico et al., 2018; Kase et al., 2017; Boatemaa et al., 2019; Nerín et al., 2016; Kase et al., 2017). Glob-

ally, foodborne diseases continue to plague populations and stand as leading contributors to both illness and mortality (Bouzemrak and Marvin, 2016; Potter et al., 2012; Pádua et al., 2019; Lüth et al., 2019). Recent estimates have identified norovirus and *Campylobacter* as the most common reason behind foodborne illnesses, while fatalities have been associated with non-typhoidal *Salmonella enterica*, *Salmonella Typhi*, *Taenia solium*, hepatitis A virus, and aflatoxin (Djekic et al., 2017; Kleter et al., 2009; Bouzemrak and Marvin, 2019).

One of the repercussions of food safety issues is the necessity for food recalls, which pose substantial economic threats to both businesses and nations alike (Deng et al., 2016). This underscores the imperative of uncovering the root causes behind these incidents and the factors contributing to contamination (Zhou et al., 2020). Cross-contamination in food and beverages is a multifaceted issue that can transpire at various stages of the food processing chain, including external raw food contamination, transportation, cleaning processes, heating, food packaging, and even during food storage. Contamination events resulting in outbreaks can manifest at any point before, during, or after food processing (Scallan and Mahon, 2012; Gupta et al., 2004).

Consequently, the pivotal task of identifying the origins of contamination or the triggers for recalls is of paramount importance (Hall et al., 2013). It is essential to gain insights into potential sources and pathways of contamination leading to foodborne outbreaks and product recalls, and to devise effective measures for prevention (Tao et al., 2020; Zhou et al., 2021; Jin et al., 2020; Marvin et al., 2017). It is worth noting that there is a dearth of substantial work in the development of computational and/or analytical models addressing these concerns (Allard et al., 2016; Moumni Abdou et al., 2019). This scarcity of research can largely be attributed to the limited availability of data concerning the contributory factors associated with food safety incidents.

As such, there is a pressing need to develop an automated tool that can mine reported food safety incidents and recalls to bridge this knowledge gap.

In this paper, we propose an event detection model that can identify food safety-related insights, recalls, and outbreaks mentioned in regulatory reports, News and social media platforms. The proposed model uses a multi-tasking sequence labeling architecture that works with transformer-based embeddings. The model leverages Liquid Time-Constant Networks (LTCs), (Hasani et al., 2021) known for their continuous-time recurrent dynamics, and integrates complex linguistic features as parameters. Tested on multiple openly available datasets, LTCs offer efficient parameter training, requiring far fewer resources than transformer models and LLMs. This is the first study to apply LTCs to multi-tasking frameworks, highlighting their efficiency compared to other time series models.

2 Related Works

Applications of language technologies and data science for food-borne risk assessment are gaining ground (Harris et al., 2017; Altenburger and Ho, 2019; Maharana et al., 2019; Deng et al., 2021). data-intensive systems play important roles in tracking food-borne illness cases and agents (Pujahari and Khan, 2022; Oldroyd et al., 2021; Nychas et al., 2021; Gupta et al., 2004; Scallan and Mahon, 2012; Wang et al., 2021). Examples at the US federal level include PulseNet (Swaminathan et al., 2001), the National Antimicrobial Resistance Monitoring System (NARMS) (Gupta et al., 2004), FoodNet (Scallan and Mahon, 2012), and the National Outbreak Reporting System (Hall et al., 2013). Implementation of whole-genome sequencing (WGS) in surveillance and outbreak investigation has fueled an explosion of publicly available foodborne pathogen genomes in new systems such as Genome-Trakr (Allard et al., 2016), EnteroBase (Zhou et al., 2020), and the National Center for Biotechnology Information’s Pathogen Detection.

These works are primarily focused on a) identifying the sentiment polarity of the documents, and b) identifying the occurrences of a set of predefined types of entities and events. However, none of the above works perform deep linguistic analysis of the textual contents to identify food-related incidents based only on the linguistic structure of the text. The model presented in this paper is distinct from the earlier approaches since it is capable of

detecting novel and heretofore unknown incidents from reports based only on semantic and linguistic analysis of content.

3 A Multi-tasking Graph Induced Liquid Time Constant Network for Food Safety Knowledge Extraction

We divide the problem statement into two tasks:

Task-1: Label sentences of each document as *Food Recall* - if the document reported events of a product recall or *Disease Outbreak* - if the document mentions events that report an foodborne disease outbreak or *NEUTRAL*- in case none of the above factors hold.

Task-2: This task had two components: (a) From among the named entities, the target organization, product name, infection, and locations were marked, if any, and (b) Mark phrases in the text that indicate food safety incidents, and its cause.

The proposed model works on each sentence at a time to detect elements of interest that are defined as follows:

1. **Target Organization (TO):** Name of the organization whose product has been recalled or found to be hazardous or caused the outbreak.
2. **Product Name (P):** The name of the product that has been recalled or caused the outbreak.
3. **Infection Name (I):** The name of the bacterial infection mentioned in the report that causes the outbreak/recall.
4. **Safety Incident (SI) -** Phrases or sequences of words that collectively are indicative of the food safety incident.
5. **Cause of Incident (CI) -** Phrases or sequences of words that indicate the primary cause of the food safety incident.
6. **Number of People Affected (N):** Phrases or sequences of words that collectively are indicative of a number of people affected due to the outbreak.

3.1 The Liquid Time Constant Network

Liquid Neural Networks were inspired by the microscopic nematode *C.elegans*, a worm that only has 302 neurons in its nervous system but still manages to respond dynamically to its environment (Hasani et al., 2021). The liquid time-constant

<p><i>recalled 36 pounds of fully cooked pork baby back ribs in</i> /SI today because they were not presented for inspection at the U.S. border. The problem was discovered when U.S. Department of Agriculture Food Safety and Inspection Service import staff reviewed records and discovered that the independent third-party carrier <i>[did not present a product for USDA inspection at the U.S.-Canadian border]</i>/CI. According to a Public Health Alert released by FSIS, being recalled are 18-pound cases containing 1.5-pound packages of "<i>[Cobblestone Farms Fully Cooked Pork Baby Back Ribs in Honey Garlic Barbeque Sauce]</i>" bearing package code "Sell By 2015-AL-08" and case code "15201" bearing the Canadian mark of inspection with establishment number "624." The product was distributed to a retailer in <i>[New York]</i>/L. In its announcement, FSIS stated that it is working on solutions to prevent future failure-to-present episodes from occurring, including outreach to industry, foreign food-safety agencies, and importers.</p>
<p><i>[ORGName]</i>/TO of Detroit, MI, is <i>[recalling approximately 1.8 million pounds]</i>/SI of <i>[ground beef products]</i>/P that may be <i>[contaminated with E. coli O157:H7]</i>/CI, the U.S. Department of Agriculture's Food Safety and Inspection Service (FSIS) announced Monday. At the time that the recall was issued, there were <i>[11 illnesses]</i>/N linked to the recalled product.</p>

Table 1: Sample Food Safety News texts with the respective annotated entities and events. Note that all the target organization names were intentionally masked by the token [ORGName] to maintain anonymity.

(LTC) model is based on neurons in the form of differential equations interconnected via sigmoidal synapses. Hasani et al. formulated the LTC model by updating the hidden state flow of continuous-time recurrent neural network (CT-RNN) (Funahashi and Nakamura, 1993). The updated hidden state flow is determined by nonlinear ODE:

$$\frac{dx(t)}{dt} = - \left[\frac{1}{\tau} + f(x(t), I(t), t, \theta) \right] x(t) + f(x(t), I(t), t, \theta)A$$

The above equation manifests a time-continuous RNN instance where $x(t)$ is the hidden state, $I(t)$ is inputs, τ is the time-constant and neural network f is parametrized by θ and A . LTC network has several features and benefits such as a) Input-dependent time-constant τ_{sys} , determined by neural network f , which adds the "liquid" nature to continuous-time RNN equations. Due to the varying nature of τ_{sys} , the network can be updated even after training is finished.

$$\tau_{sys} = \frac{\tau}{1 + \tau f(x(t), I(t), t, \theta)} \quad (1)$$

b) Solving the hidden state flow analytically is non-trivial due to the nonlinearity of the LTC state equation. As LTCs' ODE is realized by a system of stiff equations (Lepik and Hein, 2014), Hasani et al. designed a new ODE solver that fuses the explicit and the implicit Euler methods (Press, 2007) and achieves stability for an implicit update equation. The Fused Solver numerically unrolls a given dynamical system of the form $\frac{dx}{dt} = f(x)$ by:

$$x(t_{i+1}) = x(t_i) + \Delta t f(x(t_i), x(t_{i+1})) \quad (2)$$

c) LTCs realize differentiable computational graphs and can be trained by gradient-based optimization algorithms. To preserve numerical pre-

cision during a backward-pass, vanilla backpropagation through-time algorithm has been used to optimize LTCs. In 2021, Hasani et al. proved that the time-constant and the state of LTC neurons are bounded to a finite range. The *state stability* property guarantees that the outputs of LTCs never explode even if their inputs grow to infinity.

Expressivity in neural networks refers to their ability to approximate diverse functions and capture complex patterns, influenced by their architecture, depth, width, and activation functions. It has been found that LTC networks are universal approximators, meaning any autonomous ODE can be approximated by an LTC network with a finite number of neurons to any desired precision. Also, trajectory length, defined as the arc length $l(I(t)) = \int_t \left\| \frac{dI(t)}{dt} \right\| dt$, was proposed by (Raghu et al., 2017) as a measure of deep neural network expressivity. Previous experiments indicate that LTC networks exhibit greater expressivity than existing models.

To construct the Liquid Neural Network for our current model, we employ the Neural Circuit Policies (NCP) wiring (Lechner et al., 2020), specifically the AutoNCP configuration. NCPs are biologically inspired network architectures that mimic the structure of neural circuits found in living organisms. They allow for the creation of compact and efficient neural networks by leveraging sparsity and modularity. AutoNCP duplicates the nervous system of *C. elegans* nematode which consists of 4 layer of neurons namely, *Sensory Neurons*: Receives environmental information, *Inter Neurons*: Pass information gathered by sensory neurons to command neurons. The sensory neurons and inter neurons are connected by feed forward connection,

Command Neurons: Generate an output decision according to the data received from Inter Neurons. There is recurrent connection between inter neurons and command neurons and *Motor Neurons*: Execute the decision generated by the Command Neurons. There is a feed forward network between command neurons to motor neurons.

4 Augmenting GCN with BERT

In a directed graph $G = (V, E)$, where V is the set of vertices and E is the set of edges, an edge from node u to node v with label l_{uv} is denoted as (u, v, l_{uv}) . However, the flow of information along a directed edge does not always follow its direction. To account for this, and following Marcheggiani and Titov (2017), we define an updated edge set E' , which includes both the original edges in E and their inverse counterparts (v, u, l_{uv}^{-1}) .

For each node v in G , we start with an initial representation $x_v \in \mathbb{R}^d$ for all $v \in V$. Upon applying a Graph Convolutional Network (GCN), the hidden representation $h_v \in \mathbb{R}^d$ for each node is updated by considering only its immediate neighbors (Kipf and Welling, 2017). This update can be formulated as:

$$h_v = f\left(\sum_{u \in N(v)} (W_{l_{uv}}x_u + b_{l_{uv}})\right) \quad (3)$$

Here, $W_{l_{uv}} \in \mathbb{R}^{d \times d}$ and $b_{l_{uv}} \in \mathbb{R}^d$ represent label-dependent parameters that are learned during training to optimize the downstream task. The parameters $W_{l_{uv}}$ and $b_{l_{uv}}$ vary depending on the label l_{uv} of the edge connecting nodes u and v . Additionally, $N(v)$ refers to the set of neighbors of node v , which are determined based on the updated edge set E' that includes both the original and reverse edges. The function f represents a non-linear activation function, such as ReLU or sigmoid, applied element-wise to the output of the graph convolution operation.

For each sentence in the given set s_i , which consists of m tokens $\{w_1, w_2, \dots, w_m\}$, previous studies (Mintz et al., 2009; He et al., 2018) have utilized features from syntactic dependency trees to enhance relation extraction tasks. Motivated by these works, we adopt Syntactic Graph Convolution Networks (GCNs) to encode syntactic information. Specifically, for a given sentence, we first generate its syntactic dependency tree using the Stanford CoreNLP toolkit (Manning et al., 2014). This dependency tree represents syntactic relation-

ships between words, where edges represent dependencies between tokens.

Once the dependency graph is constructed, we apply a Graph Convolutional Network over this graph to capture the structural relationships between words. The dependency graph contains 55 different edge labels that represent various syntactic relations between words. Incorporating all of these edge labels would significantly overparameterize the model, which could lead to inefficient training and poor generalization. Therefore, following the approaches of Marcheggiani and Titov (2017), Nguyen and Grishman (2018), and Vashishth et al. (2018a), we simplify the dependency relations by using only two edge labels based on the direction of the syntactic dependency.

Specifically, we define a new edge label L_{uv} for an edge (u, v, l_{uv}) as follows:

- *forward*: if there is a dependency edge between two words w_1 and w_2 , and the direction of the dependency is from w_1 to w_2 .
- *backward*: if there is a dependency edge between w_1 and w_2 , and the direction of the dependency is from w_2 to w_1 .

By reducing the number of edge labels to just these two directions, we minimize the complexity of the model while still effectively capturing the syntactic structure of the sentence. This simplified approach ensures that the GCN can focus on learning the most critical syntactic relationships relevant to the downstream task, such as relation extraction, without being overwhelmed by an excessive number of dependency types.

4.1 Training the GCN augmented multi-task LTC network

Our GCN-augmented multi-task LTC (GCN+MT+LTC) network combines BERT embeddings, liquid layers, and a fully connected layer for classification and extraction tasks. It consists of three components: BERT for text embedding, a Liquid Neural Network (LTC) for sequence processing, and a fully connected layer for final classification.

We use the *WORD2VEC* model to generate embeddings for input text. This vector is passed to the LTC model. As LTC network generally perform on time-series data, to match the expected input format of the LTC model we add a randomly generated channel (time) dimension to the input sequence.

The Liquid Layer captures temporal dependencies, with 28 sensory neurons and 1 motor neuron. The output is processed by a fully connected Readout Layer, followed by ReLU and Sigmoid activation functions for binary classification and final label prediction.

The LTC network forms the core of our sequence processing component, which is critical to capture the temporal dependencies inherent in sequential data. Here, we have set the number of sensory neurons as 28 and the number of motor neurons as 1 which is same as the output size. To construct the AutoNCP circuit, we only need to specify the number of sensory and motor neurons; other neurons will be parameterized accordingly. The output of the *Liquid Layer* is passed through a fully connected *Readout Layer*. This layer maps the high-dimensional states to the output, which is then processed by subsequent layers. The output of the fully connected layer is passed through a ReLU activation function (Agarap, 2019) and a Sigmoid activation function (Kyurkchiev and Markov, 2015) for binary classification. This layer gives the final predicted label for the input sentence.

To obtain the multi-tasking LTC network for dual tasks of sequence classification and sequence labeling, the liquid layers have been trained with two separate loss functions L_1 and L_2 . Where,

$$L_1(\theta) = - \sum_{t=1}^M \sum_{k=1}^K \bar{y}_t^k \log(y_t)$$

and

$$L_2(\theta) = - \sum_{t=1}^N \sum_{j=1}^J \bar{q}_t^{i,j} \log(q_t^i).$$

Here, q_t is the vector representation of the predicted output of the model for the input word w_t^i . K and J are the number of class labels for each task. The model is fine-tuned end-to-end by minimizing the cross-entropy loss.

We define the joint loss function using a linear combination of the loss functions of the two tasks:

$$L_{joint}(\theta) = \lambda * L_1(\theta) + (1 - \lambda) * I_{[y_{sentence}=1]} * L_2(\theta)$$

Where λ controls the contribution of losses of the individual tasks in the overall joint loss. $I_{[y_{sentence}=1]}$ is an indicator function that activates the loss only when the corresponding sentence classification label is 1 since we do not want to back-propagate sequence labeling loss when the corresponding sentence classification label is 0.

5 Evaluation and Results

5.0.1 Dataset

The dataset comprises regulatory news articles from two sources namely,

- The FORCE dataset (Jana et al., 2024): The dataset is composed of a corpus of 6000 regulatory articles comprising around 121080 sentences, under Outbreak(O) and Recall(R) categories published between 2011 and 2023 by Food Safety News (FSN)¹ and a collection of around 2200 news articles from United States Food and Drug Administration (FDA) recall and outbreak announcements². All together there are annotated 8100 articles.
- SemEval 2025 Task9³: The Food Hazard Detection Dataset: The dataset for this task consists of 6,644 short texts. It includes manually labeled English food recall titles from official food agency websites (e.g., FDA).

5.0.2 Baseline models

The performance of the proposed GCN+MT+LTC model has been compared with a number of baseline models used for document classification and sequence labeling tasks such as BiLSTM, CNN-BiLSTM, RoBERTa as well as large language models like LLaMA-3 and fine-tuned Mistral 7B.

5.0.3 Experiments

Based on the given datasets, we have conducted three different experiments.

In Experiment-I: We take each of the individual datasets divide them into two groups 70% and 30% for training and testing respectively. We have performed several experiments to identify the best model architecture for our task.

In Experiment-II: We have combined both the datasets and formed a combined annotated corpus of 8844 documents. We then divide the entire corpus into 70% and 30% for training and testing respectively. The entire training set is then used to evaluate the proposed models.

In Experiment-III: To test liquid neural network based model's ability to learn from unseen data after the completion of training, we have chosen individual datasets, trained our GCN+MT+LTC model over the chosen dataset and finally tested

¹<https://www.foodsafetynews.com/>

²<https://www.fda.gov/safety>

³<https://food-hazard-detection-SemEval-2025.github.io/>

the model over other dataset. For example, we have trained our models on the *FORCE* dataset and tested the models on the *SemEval* dataset.

5.1 Results

Table 2 showcases the F1 scores for classifying food safety events into *food recall* or *disease outbreak* categories. The model achieved a peak F1 score of 0.87, coupled with a high precision of 0.89, indicating a strong ability to correctly identify relevant events with minimal false positives. Table 2 further details the accuracy of sequence labeling within sentences, categorizing words/phrases into *TO*, *P*, *I*, *SI*, *CI*, or *N* classes. This task evaluates the model’s capacity to capture complex syntactic and semantic relationships in text. The multi-objective architecture, specifically the Multi-task LTC model demonstrated superior performance across several critical categories, such as *P*, *I*, *SI*, and *CI*, outperforming the state-of-the-art models. This underscores the efficacy of leveraging multi-task learning to improve generalization and capture nuanced domain-specific patterns.

The graph augmented Multi-task LTC’s notable performance in the *P*, *I*, *SI*, and *CI* categories highlights its capacity to integrate context effectively, benefiting from its layered architecture which combines BERT’s pre-trained contextual embeddings, with liquid neural networks ability to capture contextual expressivity.

LLMs such as LLaMA-3, LLaMA 3.1 and Mistral-7B model exhibited stronger performance in identifying *CI*, and *N* classes, likely due to its fine-tuning on the dataset. Despite this, challenges in phrase boundary detection persisted, particularly for *incident* and *cause* spans, impacting its sequence labeling accuracy. However, these models are trained over billions of parameters requiring high GPU resources. On the other hand, the result table demonstrates that resource efficient models such as multi-task LTC can perform better in most of the cases as compared to the baseline models. Moreover, LLMs struggled with recognizing critical domain-specific entities such as *Salmonella*, *Listeria monocytogenes*, and *E. Coli*, indicating a lack of specialized environmental domain knowledge. This limitation significantly impacted its ability to precisely classify and label safety incidents.

We also observed that Both LLaMA and Mistral-7B often identified relevant safety incident and cause phrases but failed to accurately determine

their spans. The models frequently produced phrase lengths that were either overly broad or overly narrow, leading to penalization in sequence labeling tasks. This span detection issue highlights the challenge of capturing the precise boundaries of domain-specific phrases, which is critical for accurate labeling in complex text sequences. The results emphasize the need for enhanced domain adaptation and phrase boundary detection techniques in large language models. Incorporating domain-specific pre-training or fine-tuning, coupled with advanced span prediction strategies, could significantly improve the performance of these models in domain-intensive tasks such as food safety event classification and labeling.

Table 4 presents the outcomes of Experiment III, focusing on both classification and sequence labeling tasks at cross-domain train-test setup. In this setup, models were trained on a single dataset and evaluated on other dataset to assess their generalizability. The results are reported specifically for the BERT+LTC model, as it consistently outperformed other baseline models.

For the classification task, the GCN+MT+LTC model achieved optimal performance when trained on the SemEval dataset and tested on the FORCE dataset.

Conversely, the sequence labeling task presented more challenges, particularly when models trained on the FORCE dataset were evaluated on the other dataset. This drop in performance is likely due to the distinct linguistic characteristics and stylistic variations inherent in the FORCE dataset, which primarily consists of News data. The divergence in domain-specific language features between News texts and other dataset domains may hinder the model’s ability to effectively capture and generalize the underlying patterns required for accurate sequence labeling in different contexts.

The cross-dataset performance underscores the importance of dataset characteristics, including size, domain, and linguistic style, in influencing model generalizability. While the GCN+MT+LTC model shows promise in handling stylistic variations, the differences in domain-specific language models, particularly between News and other textual genres, pose challenges that need further exploration to enhance model adaptability across diverse datasets.

Dataset: FORCE								
No.	Model Name	SC	SL					
			TO	P	I	SI	CI	N
I.	CNN-BiLSTM	0.63	0.69	0.62	0.76	0.78	0.77	0.71
II.	RoBERTa	0.75	0.72	0.73	0.77	0.78	0.78	0.69
III.	Mistral7B	0.72	0.75	0.73	0.76	0.78	0.77	0.71
IV.	MT+LTC	0.75	0.79	0.78	0.80	0.81	0.80	0.71
V.	LLaMA-3	0.80	0.80	0.79	0.81	0.81	0.82	0.80
VI.	LLaMA3.1(8B)	0.81	0.81	0.80	0.82	0.83	0.83	0.79
VII.	GCN+MT+LTC	0.83	0.82	0.81	0.81	0.87	0.84	0.79
Dataset: SemEval Task-10								
I.	CNN-BiLSTM	0.60	0.67	0.63	0.74	0.72	0.72	0.70
II.	RoBERTa	0.76	0.70	0.71	0.75	0.72	0.72	0.69
III.	Mistral7B	0.78	0.74	0.75	0.77	0.77	0.75	0.72
IV.	MT+LTC	0.74	0.75	0.75	0.76	0.77	0.78	0.70
V.	LLaMA-3	0.79	0.81	0.73	0.79	0.78	0.75	0.73
VI.	LLaMA3.1(8B)	0.80	0.80	0.75	0.81	0.78	0.78	0.76
VII.	GCN+MT+LTC	0.82	0.80	0.77	0.84	0.82	0.80	0.78

Table 2: Results reporting F1-score of sentence classification (SC) of food safety events as *Food Recall* or *Disease Outbreak* and sequence labeling (SL) tasks.

Dataset: Combined								
No.	Model Name	SC	SL					
			TO	P	I	SI	CI	N
I.	CNN-BiLSTM	0.65	0.75	0.68	0.69	0.77	0.70	0.74
II.	RoBERTa	0.70	0.77	0.72	0.74	0.78	0.74	0.77
III.	Mistral7B	0.79	0.75	0.75	0.76	0.77	0.79	0.80
IV.	LLaMA-3	0.82	0.83	0.82	0.82	0.82	0.84	0.85
V.	MT+LTC	0.80	0.80	0.79	0.80	0.78	0.76	0.85
VI.	LLaMA3.1	0.80	0.82	0.80	0.81	0.81	0.80	0.87
VII.	GCN+MT+LTC	0.84	0.83	0.84	0.83	0.84	0.87	0.89

Table 3: Results (F1-Scores) for the combined dataset as explained in Experiment-II for the food safety incident classification (SC) and entity extraction(SL) task.

5.1.1 Domain Extensibility

LTC models demonstrate strong domain extensibility, as their temporal dynamics allow learning even after training. In Experiment-III, we trained the model on one dataset and evaluated it on another, comparing it with our previous model. Our result shows that the F1 score exceeds 85% when using SemEval Dataset as test set. However, when testing on FORCE, the Precision score drops, resulting in a lower F1 score due to the linguistic changes observed in *FORCE* dataset.

5.1.2 Performance on Highly Skewed Dataset

To assess how well the LTC model adapts to a highly skewed dataset, we have altered the combined dataset as discussed in Experiment-II. Here, we intentionally removed a large number of negative samples, thus, making the dataset intentionally skewed. In this new dataset, the count of positive and negative samples are in the ratio of 1:5 respectively. Results shows that when over 80% of the dataset is labeled as positive, traditional models like BiLSTM and BERT label all input samples

as positive and RoBERTa gives around 0.31 as F1 score. Whereas the proposed multi-task LTC model achieves an F1 score of 0.81. This finding underscores the efficacy of the LTC model in highly skewed datasets and highlights the importance of incorporating multi-task network in the architecture; we have separately experimented the LTC model using only single task (i.e either classification or entity extraction) and found an average F1 score of around 0.51 for all the entities.

5.1.3 Resource Efficiency

We found the Liquid Neural Networks as highly resource-efficient, making them ideal for CPU memory operations, unlike transformer-based models like BERT, or LLMs which demand significant GPU resources. The BERT-base model requires approximately 20.3 GB of GPU memory, while both the LTC and multi-task LTC models use less than $1.2 \times e^6$ bytes, demonstrating their minimal memory requirements compared to transformers. We also assessed LTC models against large-scale models like LLaMA-3, and LLaMA-3.1 (8 billion pa-

Sequence Labeling Task												
	FORCE (F)						SemEval (S)					
	TO	P	I	SI	CI	N	TO	P	I	SI	CI	N
F	X	X	X	X	X	X	0.82	0.86	0.82	0.78	0.79	0.78
S	0.82	0.81	0.85	0.80	0.82	0.81	X	X	X	X	X	X

Table 4: Results of Experiment-III, for the food incident extraction task, demonstrating F1 Scores for the GCN+MT+LTC model when trained over a given dataset D_i and tested over other datasets D_j such that $i \neq j$.

rameters), where both of them underperformed during inference compared to our LTC-based text classification and extraction model, which uses only 100K parameters and 28 neurons in the AutoNCP wiring. Training large language models is resource- and time-intensive, while training the LTC model is much faster. For instance, LTC trains for 50 epochs on domain-specific datasets in just 1 hour. The training time per epoch for all the datasets is negligible, with prediction times per sample taking only milliseconds.

6 Conclusion

In this paper we present automatic extraction of critical information related to food safety and foodborne infection from regulatory reports and social media platforms. The proposed model, founded on a multi-tasking sequence labeling architecture integrated with a graph enabled liquid time-constant neural network, demonstrates its effectiveness not only in terms of accuracy but also shows promising results in terms of domain extensibility, data-skewedness and resource efficiency. To develop and evaluate our model, we meticulously curated an annotated corpus comprising pertinent articles. Our initial analysis demonstrate the proposed multi-task model surpasses the performance of almost all the baseline models including LLMs such as LLaMA-3 and fine-tuned mistral-7B.

7 Limitations

Although the proposed multi-tasking sequence labeling model integrated with a graph-enabled liquid time-constant network shows strong performance, several limitations exist: a) Domain Coverage: The model has been evaluated primarily on food safety and food-borne infection data. Its adaptability to other regulatory or health-related domains remains to be validated. b) Reliance on Curated Corpus: Performance depends on the quality and representativeness of the annotated corpus. In real-world scenarios with unstructured or noisy data,

accuracy may vary. c) Interpretability: Despite improved accuracy, the underlying reasoning process of the graph-enabled LTC network is complex, limiting transparency for critical decision-making. d) Scalability: While resource-efficient, handling extremely large-scale social media streams or multilingual data may pose challenges. e) Knowledge Gaps: The model may struggle when encountering emerging food safety issues or novel infection patterns not represented in the training data.

8 Ethical Considerations

This work relies exclusively on previously published, publicly available datasets and pretrained language models for food recall and foodborne outbreak detection. The primary data sources consist of regulatory announcements and news articles released by official agencies (e.g., food safety authorities) and reputable media outlets. No data was collected directly from individuals, and no personally identifiable information (PII) was intentionally introduced during dataset creation or model training.

References

- Abien Fred Agarap. 2019. [Deep learning using rectified linear units \(relu\)](#). *Preprint*, arXiv:1803.08375.
- Marc W Allard, Errol Strain, David Melka, Kelly Bunning, Steven M Musser, Eric W Brown, and Ruth Timme. 2016. Practical value of food pathogen traceability through building a whole-genome sequencing network and database. *Journal of clinical microbiology*, 54(8):1975–1983.
- Kristen M Altenburger and Daniel E Ho. 2019. Is yelp actually cleaning up the restaurant industry? a re-analysis on the relative usefulness of consumer reviews. In *The World Wide Web Conference*, pages 2543–2550.
- Priscilla D’ Amico, Daniele Nucera, Lisa Guardone, Martino Mariotti, Roberta Nuvoloni, and Andrea Armani. 2018. Seafood products notifications in the eu rapid alert system for food and feed (rasff) database:

- Data analysis during the period 2011–2015. *Food Control*, 93:241–250.
- Sandra Boatemaa, McKenna Barney, Scott Drimie, Julia Harper, Lise Korsten, and Laura Pereira. 2019. Awakening from the listeriosis crisis: Food safety challenges, practices and governance in the food retail sector in south africa. *Food Control*, 104:333–342.
- Yamine Bouzembrak and Hans JP Marvin. 2016. Prediction of food fraud type using data from rapid alert system for food and feed (rasff) and bayesian network modelling. *Food Control*, 61:180–187.
- Yamine Bouzembrak and Hans JP Marvin. 2019. Impact of drivers of change, including climatic factors, on the occurrence of chemical food safety hazards in fruits and vegetables: A bayesian network approach. *Food control*, 97:67–76.
- Xiangyu Deng, Shuhao Cao, and Abigail L Horn. 2021. Emerging applications of machine learning in food safety. *Annual Review of Food Science and Technology*, 12:513–538.
- Xiangyu Deng, Henk C den Bakker, and Rene S Hendriksen. 2016. Genomic epidemiology: whole-genome-sequencing-powered surveillance and outbreak investigation of foodborne bacterial pathogens. *Annual review of food science and technology*, 7:353–374.
- Ilija Djekic, Danijela Jankovic, and Andreja Rajkovic. 2017. Analysis of foreign bodies present in european food using data from rapid alert system for food and feed (rasff). *Food control*, 79:143–149.
- Ken-ichi Funahashi and Yuichi Nakamura. 1993. Approximation of dynamical systems by continuous time recurrent neural networks. *Neural networks*, 6(6):801–806.
- Amita Gupta, Jennifer M Nelson, Timothy J Barrett, Robert V Tauxe, Shannon P Rossiter, Cindy R Friedman, Kevin W Joyce, Kirk E Smith, Timothy F Jones, Marguerite A Hawkins, and 1 others. 2004. Antimicrobial resistance among campylobacter strains, united states, 1997–2001. *Emerging infectious diseases*, 10(6):1102.
- Aron J Hall, Mary E Wikswa, Karunya Manikonda, Virginia A Roberts, Jonathan S Yoder, and L Hannah Gould. 2013. Acute gastroenteritis surveillance through the national outbreak reporting system, united states. *Emerging infectious diseases*, 19(8):1305.
- Jenine K Harris, Jared B Hawkins, Leila Nguyen, Elaine O Nsoesie, Gaurav Tuli, Raed Mansour, and John S Brownstein. 2017. Research brief report: Using twitter to identify and respond to food poisoning: The food safety stl project. *Journal of Public Health Management and Practice*, 23(6):577.
- Ramin Hasani, Mathias Lechner, Alexander Amini, Daniela Rus, and Radu Grosu. 2021. Liquid time-constant networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 7657–7666.
- Sudeshna Jana, Manjira Sinha, and Tirthankar Dasgupta. 2024. Force: A benchmark dataset for foodborne disease outbreak and recall event extraction from news. In *Proceedings of The 9th Social Media Mining for Health Research and Applications (SMM4H 2024) Workshop and Shared Tasks*, pages 163–169.
- Cangyu Jin, Yamine Bouzembrak, Jiehong Zhou, Qiao Liang, Leonieke M Van Den Bulk, Anand Gavai, Ningjing Liu, Lukas J Van Den Heuvel, Wouter Hoenderdaal, and Hans JP Marvin. 2020. Big data in food safety—a review. *Current Opinion in Food Science*, 36:24–32.
- Julie Ann Kase, Guodong Zhang, and Yi Chen. 2017. Recent foodborne outbreaks in the united states linked to atypical vehicles—lessons learned. *Current Opinion in Food Science*, 18:56–63.
- GA Kleter, ALDO Prandini, LAURA Filippi, and HJP Marvin. 2009. Identification of potentially emerging food safety issues by analysis of reports published by the european community’s rapid alert system for food and feed (rasff) during a four-year period. *Food and chemical toxicology*, 47(5):932–950.
- Nikolay Kyurkchiev and Svetoslav Markov. 2015. Sigmoid functions: some approximation and modelling aspects. *LAP LAMBERT Academic Publishing, Saarbrücken*, 4.
- Mathias Lechner, Ramin Hasani, Alexander Amini, Thomas A Henzinger, Daniela Rus, and Radu Grosu. 2020. Neural circuit policies enabling auditable autonomy. *Nature Machine Intelligence*, 2(10):642–652.
- Ülo Lepik and Helle Hein. 2014. *Stiff Equations*, pages 45–57.
- Stefanie Lüth, Idesbald Boone, Sylvia Kleta, and Sascha Al Dahouk. 2019. Analysis of rasff notifications on food products contaminated with listeria monocytogenes reveals options for improvement in the rapid alert system for food and feed. *Food Control*, 96:479–487.
- Adyasha Maharana, Kunlin Cai, Joseph Hellerstein, Yulin Hswen, Michael Munsell, Valentina Staneva, Miki Verma, Cynthia Vint, Derry Wijaya, and Elaine O Nsoesie. 2019. Detecting reports of unsafe foods in consumer product reviews. *JAMIA open*, 2(3):330–338.
- Hans JP Marvin, Esmée M Janssen, Yamine Bouzembrak, Peter JM Hendriksen, and Martijn Staats. 2017. Big data in food safety: An overview. *Critical reviews in food science and nutrition*, 57(11):2286–2295.

- Houda Moumni Abdou, Ilham Dahbi, Mohammed Akrim, Fatima Zahra Meski, Yousef Khader, Mohammed Lakranbi, Hind Ezzine, and Asmae Khat-tabi. 2019. Outbreak investigation of a multipathogen foodborne disease in a training institute in rabat, morocco: case-control study. *JMIR public health and surveillance*, 5(3):e14227.
- Cristina Nerín, Margarita Aznar, and Daniel Carrizo. 2016. Food contamination during food process. *Trends in food science & technology*, 48:63–68.
- George-John Nychas, Emma Sims, Panagiotis Tsakanikas, and Fady Mohareb. 2021. Data science in the food industry. *Annual Review of Biomedical Data Science*, 4:341–367.
- Rachel A Oldroyd, Michelle A Morris, and Mark Birkin. 2021. Predicting food safety compliance for informed food outlet inspections: a machine learning approach. *International Journal of Environmental Research and Public Health*, 18(23):12635.
- Inês Pádua, André Moreira, Pedro Moreira, Filipa Melo de Vasconcelos, and Renata Barros. 2019. Impact of the regulation (eu) 1169/2011: Allergen-related recalls in the rapid alert system for food and feed (rasff) portal. *Food control*, 98:389–398.
- Antony Potter, Jason Murray, Benn Lawson, and Stephanie Graham. 2012. Trends in product recalls within the agri-food industry: Empirical evidence from the usa, uk and the republic of ireland. *Trends in food science & technology*, 28(2):77–86.
- William H Press. 2007. *Numerical recipes 3rd edition: The art of scientific computing*. Cambridge university press.
- Rakesh Mohan Pujahari and Rijwan Khan. 2022. Applications of machine learning in food safety. In *Artificial Intelligence Applications in Agriculture and Food Quality Improvement*, pages 216–240. IGI Global.
- Maithra Raghu, Ben Poole, Jon Kleinberg, Surya Ganguli, and Jascha Sohl-Dickstein. 2017. [On the expressive power of deep neural networks](#). *Preprint*, arXiv:1606.05336.
- Elaine Scallan and Barbara E Mahon. 2012. Foodborne diseases active surveillance network (foodnet) in 2012: a foundation for food safety in the united states. *Clinical Infectious Diseases*, 54(suppl_5):S381–S384.
- Bala Swaminathan, Timothy J Barrett, Susan B Hunter, Robert V Tauxe, and CDC PulseNet Task Force. 2001. Pulsenet: the molecular subtyping network for foodborne bacterial disease surveillance, united states. *Emerging infectious diseases*, 7(3):382.
- Dandan Tao, Pengkun Yang, and Hao Feng. 2020. Utilization of text mining as a big data analysis tool for food science and nutrition. *Comprehensive reviews in food science and food safety*, 19(2):875–894.
- Hanxue Wang, Wenjuan Cui, Yunchang Guo, Yi Du, Yuanchun Zhou, and 1 others. 2021. Machine learning prediction of foodborne disease pathogens: Algorithm development and validation study. *JMIR medical informatics*, 9(1):e24924.
- Qinqin Zhou, Hao Zhang, and Suya Wang. 2021. Artificial intelligence, big data, and blockchain in food safety. *International journal of food engineering*, 18(1):1–14.
- Zhemín Zhou, Nabil-Fareed Alikhan, Khaled Mohamed, Yulei Fan, Mark Achtman, Derek Brown, Marie Chataway, Tim Dallman, Richard Delahay, Christian Kornschober, and 1 others. 2020. The enterobase user’s guide, with case studies on salmonella transmissions, yersinia pestis phylogeny, and escherichia core genomic diversity. *Genome research*, 30(1):138–152.