



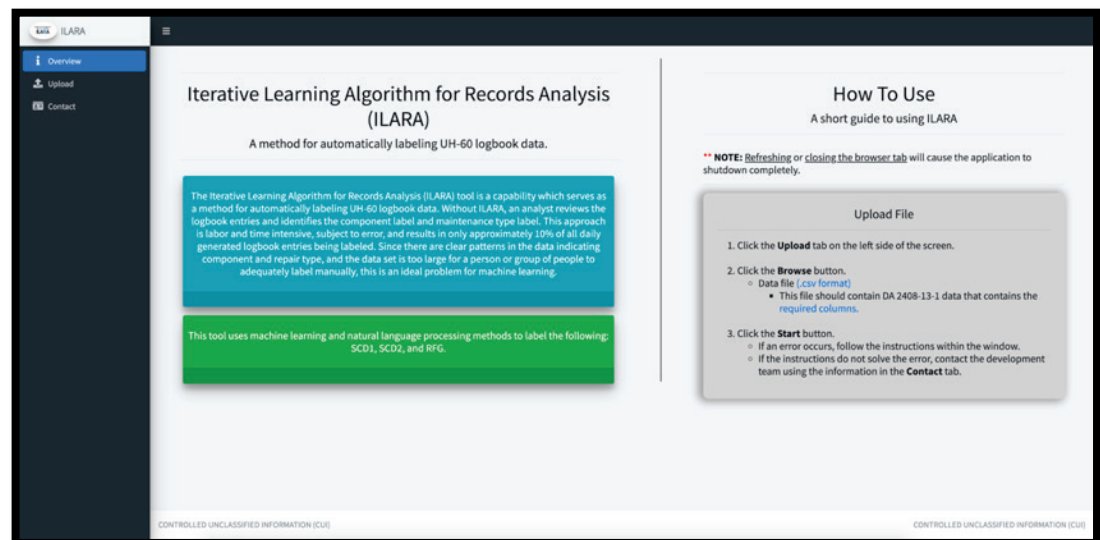
**US Army Corps  
of Engineers®**  
Engineer Research and  
Development Center



# Iterative Learning Algorithm for Records Analysis (ILARA) User Manual

Joshua Church, LaKenya Walker, and Amy Bednar

August 2021



**The U.S. Army Engineer Research and Development Center (ERDC)** solves the nation's toughest engineering and environmental challenges. ERDC develops innovative solutions in civil and military engineering, geospatial sciences, water resources, and environmental sciences for the Army, the Department of Defense, civilian agencies, and our nation's public good. Find out more at [www.erdclibrary.on.worldcat.org/v2](http://www.erdclibrary.on.worldcat.org/v2).

To search for other technical reports published by ERDC, visit the ERDC online library at <http://www.erdclibrary.on.worldcat.org/v2>.

# **Iterative Learning Algorithm for Records Analysis (ILARA) User Manual**

Joshua Church, LaKenya Walker, and Dr. Amy Bednar

*Information Technology Laboratory  
US Army Engineer Research and Development Center  
3909 Halls Ferry Road  
Vicksburg, MS 39180-6199*

Final Report

Approved for public release; distribution is unlimited.



Prepared for Joint Artificial Intelligence Center  
Fort Meade, MD 20755

Under MIPR HC1085015834

## Abstract

This manual is intended for new users with minimal or no experience with using the Iterative Learning Algorithm for Records Analysis (ILARA) tool. The goal of this document is to give an overview of the main functions of ILARA. The primary focus of this document is to demonstrate functionality.

Every effort has been made to ensure this document is an accurate representation of the functionality of the ILARA tool. For additional information about this manual, contact [ERDC.JAIC@erdc.dren.mil](mailto:ERDC.JAIC@erdc.dren.mil).

**DISCLAIMER:** The contents of this report are not to be used for advertising, publication, or promotional purposes. Citation of trade names does not constitute an official endorsement or approval of the use of such commercial products. All product names and trademarks cited are the property of their respective owners. The findings of this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.  
**DESTROY THIS REPORT WHEN NO LONGER NEEDED. DO NOT RETURN IT TO THE ORIGINATOR.**

# Contents

<b>Abstract</b> .....	<b>ii</b>
<b>Contents</b> .....	<b>iii</b>
<b>Figures</b> .....	<b>iv</b>
<b>Preface</b> .....	<b>v</b>
<b>Abbreviations and Acronyms</b> .....	<b>vi</b>
<b>1 Introduction</b> .....	<b>1</b>
1.1 Background.....	1
1.2 Purpose .....	2
1.3 Users.....	2
1.3.1 <i>Maintainer</i> .....	2
1.3.2 <i>Analyst</i> .....	2
1.4 Requirements.....	2
1.4.1 <i>Recommended operating systems</i> .....	2
1.4.2 <i>Hardware</i> .....	3
1.4.3 <i>Software</i> .....	3
1.4.4 <i>Data</i> .....	3
1.4.5 <i>Columns (logbook fields)</i> .....	4
<b>2 User Interface</b> .....	<b>5</b>
2.1 Main page .....	5
2.2 Tabs .....	5
2.2.1 <i>Overview</i> .....	6
2.2.2 <i>Upload</i> .....	6
2.3 Contact .....	10
<b>3 Summary</b> .....	<b>12</b>
<b>REPORT DOCUMENTATION PAGE</b> .....	<b>13</b>

## Figures

1	Main page window .....	5
2	Detail of tabs.....	5
3	Overview window .....	6
4	Upload file window .....	7
5	ILARA customization window.....	8
6	Labeling complete window .....	9
7	Top 25 rows of labeled data button.....	10
8	Labeled data top rows (example).....	10
9	Contact information window.....	11

## Preface

This study was conducted for the Joint Artificial Intelligence Center (JAIC) under MIPR HC1085015834 for Project “ERDC PMx Product Transition to the Joint Artificial Intelligence Center (JAIC) Joint Common Foundation (JCF).” The technical monitor was Dr. Maria Niki Goerger.

The work was performed by the Computational Analysis Branch of the Computational Science and Engineering Division, U.S. Army Engineer Research and Development Center, Information Technology Laboratory (ERDC-ITL). At the time of publication, Mr. Joshua Church was Acting Branch Chief; Dr. Jeffrey Hensley was Division Chief; and Dr. Robert Wallace was Technical Director for the Engineered Resilient Systems (ERS) program. The Deputy Director of ERDC-ITL was Dr. Jackie Pettway, and the Director was Dr. David Horner.

COL Teresa A. Schlosser was the Commander of ERDC, and Dr. David W. Pittman was the Director.

## Abbreviations and Acronyms

Acronym	Definition
AvMC	Army Aviation and Missile Center
DA	Department of the Army
DRF	Distributed Random Forest
ERS	Engineered Resilient Systems
ERDC	Engineer Research and Development Center
ILARA	Iterative Learning Algorithm for Records Analysis
ITL	Information Technology Laboratory
JAIC	Joint Artificial Intelligence Center
JCF	Joint Common Foundation
LTS	Long Term Support
ML	Machine Learning
NLP	Natural Language Processing
NMI	National Mission Initiative
PMx	Predictive Maintenance

# 1 Introduction

## 1.1 Background

The U.S. Army Engineer Research and Development Center (ERDC) Information Technology Laboratory (ITL) has several years of experience in collecting, curating, and making maintenance logbook data readily available for data analysis and reporting purposes. Specifically, ERDC-ITL has been successful in performing these tasks for maintenance and operational H-60 data across the services through projects with the Army Aviation and Missile Center (AvMC) and the Joint Artificial Intelligence Center (JAIC) Predictive Maintenance (PMx) National Mission Initiative (NMI).

ERDC-ITL developed the Iterative Learning Algorithm for Records Analysis (ILARA), a set of Natural Language Processing (NLP) and Distributed Random Forest (DRF) models that perform label propagation for maintenance logbook data. Label propagation is the process of using machine learning (ML) techniques to learn labeling patterns from existing data that has been reliably labeled by a subject matter expert. These learned patterns are then used to assign labels to new, unlabeled data.

ILARA is responsible for labeling maintenance logbook entries with (1) the component being repaired and (2) the type of maintenance being performed. These are pieces of information that are not explicitly provided by the maintainer in the field who is generating the logbook entry. Nonetheless, the component and repair type can be deduced from other data fields that are provided by the maintainer.

Currently, without ILARA, an analyst reviews the logbook entries and identifies the component label and maintenance type label. This approach is labor and time intensive, subject to error, and results in only approximately 10% of all daily generated logbook entries being labeled. Since there are clear patterns in the data indicating component and repair type, and the data set is too large for a person or group of people to adequately label manually, this is an ideal problem for ML.

The data generated by ILARA can provide awareness of the platform's maintenance landscape, revealing trends and patterns of maintenance, such as regularity of certain kinds of problems, scheduling patterns of

maintenance types, or logistical inefficiencies such as delays due to part unavailability. Revealing these characteristics of the maintenance and repair process provides opportunities for correction, optimization, and improvement.

## 1.2 Purpose

ILARA should be used by maintainers (not in the field) and analysts to aid in labeling maintenance logbook entries that are missing critical data pieces such as the component being repaired and the type of maintenance being performed. These are pieces of information that are not explicitly provided by the maintainer in the field who is generating the logbook entry. The specific targets that ILARA will label are: *SCD1* (Potential Termination), *SCD2* (Failure Classification), and *RFG* (Aircraft model designation series).

## 1.3 Users

This section discusses the different types of users.

### 1.3.1 Maintainer

A maintainer should utilize ILARA to automatically label the following fields: *SCD1*, *SCD2*, and *RFG*. This user manual will cover the steps for using ILARA.

### 1.3.2 Analyst

An analyst should utilize the output from ILARA. This data will now have automatically labeled fields (i.e., *SCD1*, *SCD2*, and *RFG*), which can be analyzed to find trends in the maintenance logbook data over specific lengths of time.

## 1.4 Requirements

This section covers the following requirements to use ILARA: *operating system, hardware, software, and data requirements*.

### 1.4.1 Recommended operating systems

- *Windows*: Windows 10
- *Mac*: macOS Sierra or newer

- *Linux*: Ubuntu  $\geq$  16.04 Long Term Support (LTS).

#### **1.4.2 Hardware**

It is recommended to have a computer less than five years old that has the following:

- Processor: Minimum 2 GHz; Recommended 3 GHz
- Hard Drive: Minimum 32 GB; Recommended 64 GB
- Memory (RAM): Minimum 8 GB; Recommended 64 GB.

#### **1.4.3 Software**

Supported browsers are

- Google Chrome, and
- Firefox.

Other important software includes

- R,
- RStudio, and
- Java 8, 9, 10, 11, 12, or 13.

#### **1.4.4 Data**

ILARA is designed to work with Department of the Army (DA) 2408-13-1 maintenance logbook data. The DA 2408-13-1 form contains aircraft inspection and maintenance records. The data in these records include

- Operational remarks and faults found during flight or operation, pre-flight, through flight, and post flight inspections;
- Checks, services, and scheduled or unscheduled maintenance inspections;
- Faults found, including: battle damage assessment and repair, when faults were found, when faults occurred, how faults were recognized, effect faults had on the operation or mission, and corrective actions taken to correct faults;
- Person-hours it takes to do maintenance and quality control work by military, civilian, and contract maintenance support personnel; and
- Condition status symbols for aircraft and mission-related equipment.

**NOTE:** This data file must be in comma-separated variables (.csv) format. For additional information, refer to the JAIC Joint Logistics/Predictive Maintenance H-60 Data Documentation tool.

#### **1.4.5 Columns (logbook fields)**

The following columns must exist within the data file:

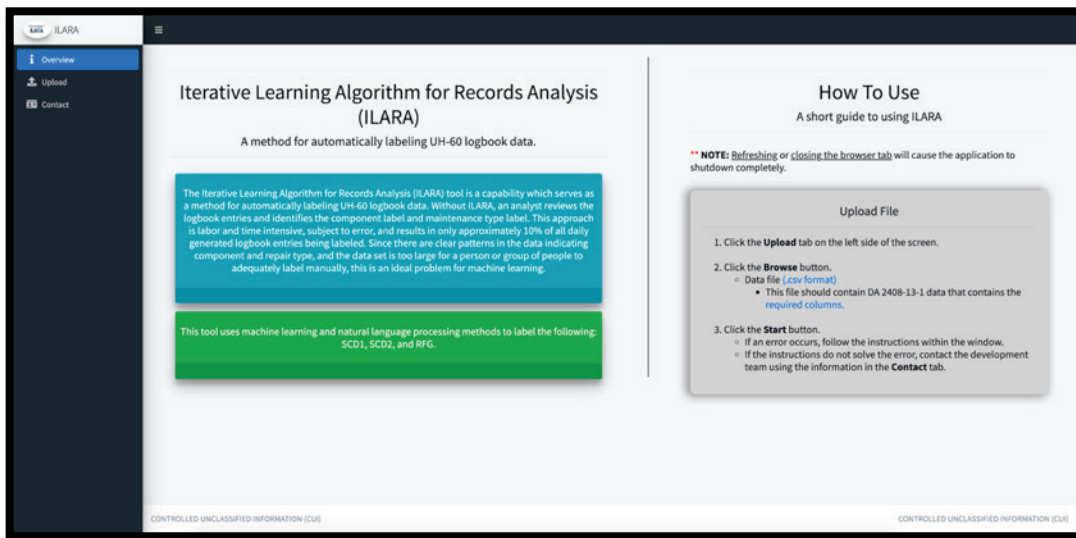
- **KEY13:** Unique identifier for each record
- **NARR:** 2408-13-1 write-up narrative
- **CORR\_NARR:** 2408-13-1 write-up for corrective narrative
- **EVENT\_DATE\_TIME:** Date the event occurred or was entered into the logbook
- **EI\_SN:** End Item Serial Number/Aircraft Tail Number
- **STATUS:** Aircraft status (e.g., grounded)
- **HOW\_REC:** How it was recognized
- **MAL\_EFF:** Malfunction effect code
- **WHEN\_DISC:** When discovered code DA PAM 738-751, Table 1-4
- **SYS\_CODE:** System Codes (e.g., Aircraft, Armament, Electronic, Other)
- **EI\_BEG\_AGE:** End Item Beginning Age (Airframe Hours)
- **TMMH:** Total maintenance man hours
- **SCD1:** Potential Termination (will be created for the user)
- **SCD2:** Failure Classification (will be created for the user)
- **RFG:** Aircraft model designation series (will be created for the user).

## 2 User Interface

### 2.1 Main page

Once the ILARA tool loads, the user is presented with the interface in Figure 1:

Figure 1. Main page window.

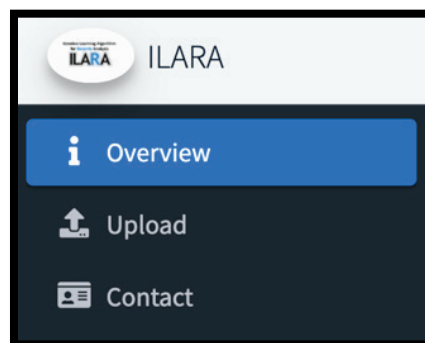


On the left side of the screen, the user can select different tab options.

### 2.2 Tabs

There are three tabs available for the user to select: **Overview**, **Upload**, and **Contact** (see Figure 2). Information about each tab can be found in the following sections.

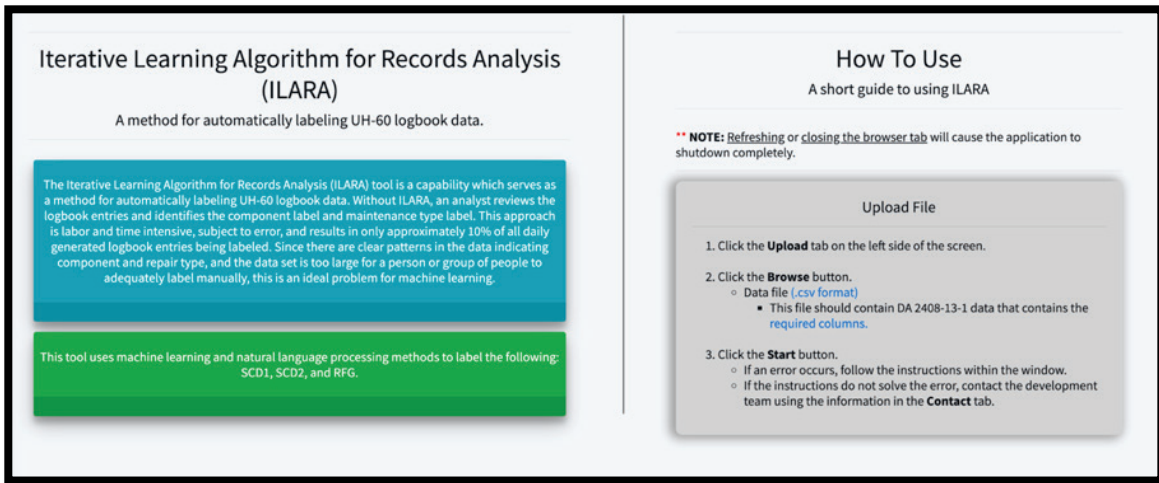
Figure 2. Detail of tabs.



### 2.2.1 Overview

The **Overview** window (see Figure 3) contains information on the purpose of ILARA and brief instructions on how to use the software. These instructions focus on how to upload a file which is in the **Upload** section.

Figure 3. Overview window.



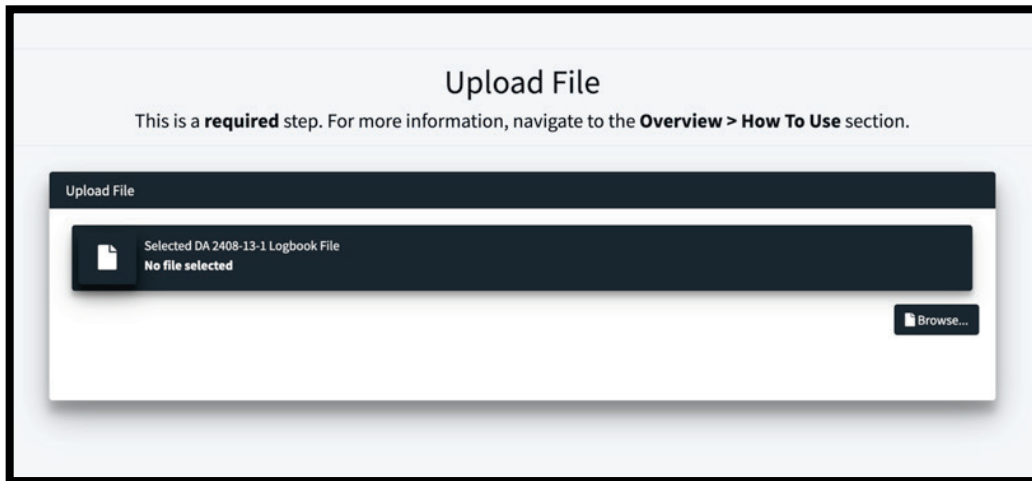
**NOTE:** When using the ILARA tool, refreshing or closing the browser tab will cause the application to shut down.

### 2.2.2 Upload

The **Upload** tab allows the user to upload a data file. This file must meet the requirements in the Data and Columns (logbook fields) sections above.

After clicking the **Upload** tab, the user will be presented with an option to upload a DA 2408-13-1 logbook data file, as shown in Figure 4.

Figure 4. Upload file window.



To upload a file, perform the following steps:

1. Click Browse...
2. Navigate to the system path that contains the DA 2408-13-1 logbook file.
3. Once selected, the ILARA Customization window will appear (see Figure 5).

**NOTE:** If one of the required columns is missing, an error window will appear. This window will provide the user with a list of the missing columns.

Figure 5. ILARA customization window.

ILARA Customization

Select the directory where to store the labeled data.

Output Directory: /Users/rditljqc/Datasets/example

Show Probability? (Optional)

If selected, this option will append the probability value for the label (i.e., the confidence value).

Keep SCD1\_Binary Column? (Optional)

If selected, this option will keep the SCD1\_Binary column. This column is used as an intermediate step to help with the labeling process; however, it is not needed once the labeling process is complete.

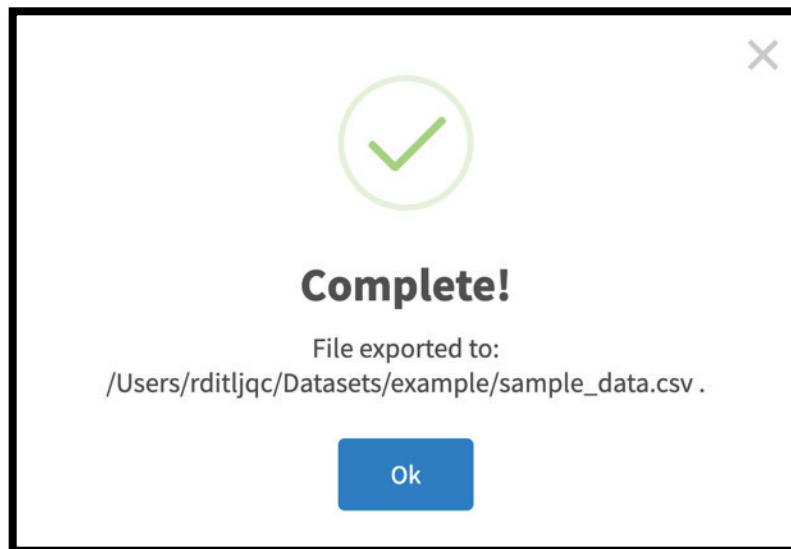
Keep Vectorized Columns? (Optional)

If selected, this option will keep the vectorized columns. These columns are the vectorized versions of the Narrative and Corrective Narrative columns. These are used as part of the labeling process; however, they are not needed after the labeling process is complete.

4. In the *ILARA Customization* window, the user will be presented with the following options:
  - a. *Output Directory (Required)*
    - (1) The output directory is where the labeled file will be exported.
  - b. *Show Probability (Optional)*
    - (1) If selected, this option will append the probability value for the label (i.e., the “confidence” percentage for why the specific label was chosen by the model).
  - c. *Keep SCD1\_Binary (Optional)*
    - (1) If selected, this option will keep the SCD1\_Binary column. This column is used as an intermediate step to help with the labeling process; however, it is not needed once the labeling process is complete.
  - d. *Keep Vectorized Columns (Optional)*
    - (1) If selected, this option will keep the vectorized columns. These columns are the vectorized versions of the Narrative and Corrective Narrative columns. These are used as part of the labeling process; however, they are not needed after the labeling process is complete.
5. Click *Submit*.
  - a. The aforementioned parameters will now be set for the ILARA labeling workflow.
6. Click *Start*.

- a. The ILARA labeling workflow will start executing. The operation time varies due to
  - (1) Data size
  - (2) Processing capabilities of the machine (i.e., number of CPU threads, etc.)
- b. The labeling workflow is broken down into seven stages:
  - (1) *Stage one* cleans the file based on defined rules.
  - (2) *Stage two* creates word embeddings from the free text columns.
  - (3) *Stage three* creates SCD1\_Binary column.
  - (4) *Stage four* labels the SCD2 column.
  - (5) *Stage five* labels the SCD1 column.
  - (6) *Stage six* labels the RFG column.
    - (a) This is the longest step in the process.
  - (7) *Stage seven* finalizes the files based on the user-defined options from the *ILARA Customization* window.
7. Once the labeling process is complete, the following window shown in Figure 6 will appear:

Figure 6. Labeling complete window.



8. Click *OK* to close the window.
  - a. A new option to view the top 25 rows is now available.
9. Click *View Top 25 Rows of the Labeled Data* to view a sample of the newly-labeled data file (see Figure 7 and Figure 8).

Figure 7. Top 25 rows of labeled data button.

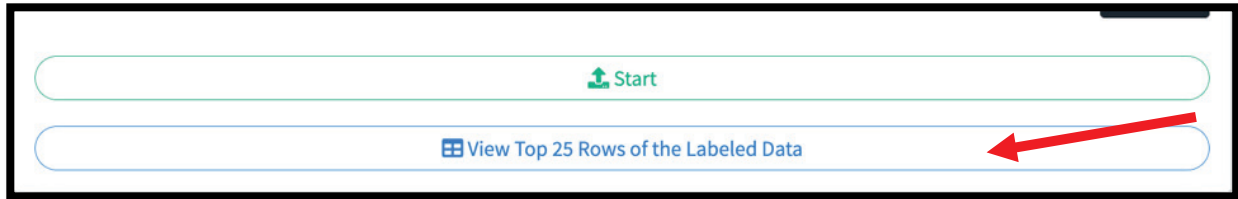


Figure 8. Labeled data top rows (example).

Labeled Data

IEN	TIMH	SYS_CODE	WHEN_DISC	CNAME	RFG	SCD1	SCD2
1		A	X	T/R GB	02_PMI2	S	N
1	0.200000002980232	A	O	Air Inlet	02_PMI2	S	N
1	0.200000002980232	A	X	Accumulator, #3 Hyd Sys	07B11	E	B
2		A	K	Fwd Fuselage Door Installation	02_360HR	N	N
1		A	O	40 Hour Inspection	02_40HR	S	N

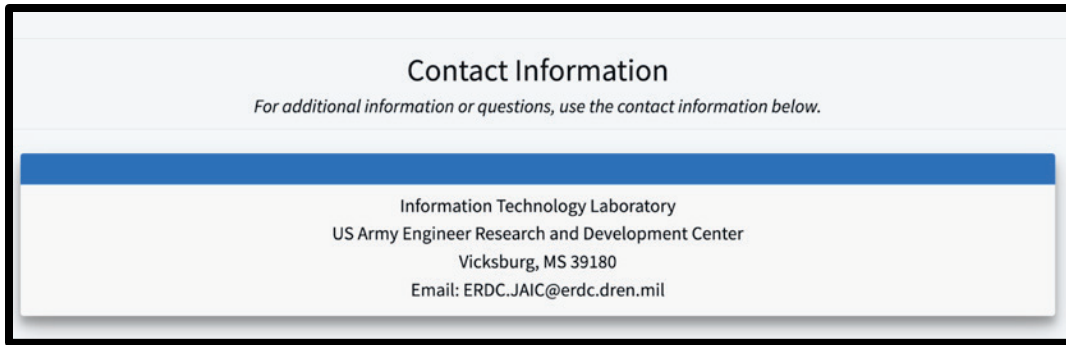
Previous 1 2 3 Next

Dismiss

### 2.3 Contact

The **Contact** window, as shown in Figure 9, provides the contact information for the developer of the ILARA tool.

Figure 9. Contact information window.



### **3 Summary**

ILARA is an application that provides an algorithm that uses machine learning techniques to learn labeling patterns from manually labeled data in order to assign labels to new data. The ultimate benefit of this tool is the ability to automate the scoring of maintenance records, drastically reducing the time and labor to perform this process.

# REPORT DOCUMENTATION PAGE

Form Approved  
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

<b>1. REPORT DATE (DD-MM-YYYY)</b> August 2021		<b>2. REPORT TYPE</b> Final Report		<b>3. DATES COVERED (From - To)</b>	
Iterative Learning Algorithm for Records Analysis (ILARA) User Manual				<b>5a. CONTRACT NUMBER</b>	
				<b>5b. GRANT NUMBER</b>	
				<b>5c. PROGRAM ELEMENT NUMBER</b>	
Joshua Church, LaKenya Walker, and Dr. Amy Bednar				<b>5d. PROJECT NUMBER</b>	
				<b>5e. TASK NUMBER</b>	
				<b>5f. WORK UNIT NUMBER</b>	
<b>7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)</b>  Information Technology Laboratory US Army Engineer Research and Development Center 3909 Halls Ferry Road Vicksburg, MS 39180-6199				<b>8. PERFORMING ORGANIZATION REPORT NUMBER</b>  ERDC/ITL SR-21-8	
<b>9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)</b>  Joint Artificial Intelligence Center Defense Information Systems Agency DISA General Fund Division 6910 Cooper Ave. Fort Meade, MD 20755				<b>10. SPONSOR/MONITOR'S ACRONYM(S)</b>  JAIC	
				<b>11. SPONSOR/MONITOR'S REPORT NUMBER(S)</b>	
<b>12. DISTRIBUTION / AVAILABILITY STATEMENT</b> Approved for public release; distribution is unlimited.					
<b>13. SUPPLEMENTARY NOTES</b> MIPR HC1085015834					
<b>14. ABSTRACT</b> This manual is intended for new users with minimal or no experience with using the Iterative Learning Algorithm for Records Analysis (ILARA) tool. The goal of this document is to give an overview of the main functions of ILARA. The primary focus of this document is to demonstrate functionality. Every effort has been made to ensure this document is an accurate representation of the functionality of the ILARA tool. For additional information about this manual, contact ERDC.JAIC@erd.c.dren.mil.					
<b>15. SUBJECT TERMS</b> Maintenance--Records--Analysis, Data processing, Natural language processing (Computer science), Machine learning					
<b>16. SECURITY CLASSIFICATION OF:</b>			<b>17. LIMITATION OF ABSTRACT</b>	<b>18. NUMBER OF PAGES</b>	<b>19a. NAME OF RESPONSIBLE PERSON</b>
<b>a. REPORT</b>	<b>b. ABSTRACT</b>	<b>c. THIS PAGE</b>			<b>19b. TELEPHONE NUMBER (include area code)</b>
Unclassified	Unclassified	Unclassified	SAR	21	